
DOMINE
UT VIDEAM

[illegible]
[illegible]

INTERMEDIATE PHYSICS

INTERMEDIATE PHYSICS

BY

W. WATSON, A.R.C.S., D.Sc. (London), F.R.S.
ASSISTANT PROFESSOR OF PHYSICS AT THE ROYAL
COLLEGE OF SCIENCE, LONDON

WITH DIAGRAMS

LONGMANS, GREEN AND CO.
39 PATERNOSTER ROW, LONDON
NEW YORK, BOMBAY, AND CALCUTTA
1912

PREFACE

When arranging a course of instruction for intermediate students, most of whom will not proceed with the study of Physics beyond this stage, one is met with the difficulty that the range of subjects which may be dealt with is so great. This difficulty is continually increasing since the applications of physical principles to engineering and the like are continually being extended, as for example in the case of wireless-telegraphy, and at the same time new branches of Physics, such as radio-activity, are being developed. In either case the student rightly expects that even in an intermediate course such subjects should not be ignored. The result of such extensions is that the teacher has either to attempt to cover all the ground, and hence owing to limitations of time can only deal very superficially with the various subjects, or he has to limit the number of subjects which he will consider. The author believes that this latter alternative is the better one, and that a considerable number of the simpler physical phenomena which it has been usual to include in an intermediate course can with advantage be omitted, or at any rate dealt with very briefly; the time thus saved being employed in examining a little more in detail some of the other branches. In this book he has attempted to carry out this idea, so that although it will be found that the number of physical phenomena considered is not quite as large as is often the case in works of a similar scope, yet some branches are carried a little beyond what is generally considered the intermediate stage. In selecting the parts to be extended in this way the author has had in view the requirements of engineering, medical and other students, who will, to a certain extent, have to apply physical principles in their subsequent studies. He feels that in this way the gap which at present is apt to exist between the course in Physics taken, say, by an engineering student, and the applications of physical principles he meets with in his engineering reading will be reduced, and at the same time the interest of the student will be better maintained than when he is introduced to a large mass of physical phenomena which, when only examined in a very elementary manner, must appear more or less disconnected. The pages on mechanics do not profess to give a complete account of the subject, only those problems being included which are required when dealing with the more physical problems considered in the main part of the book.

CONTENTS

BOOK I

MECHANICS AND PROPERTIES OF MATTER

CHAPTER I

DISPLACEMENT

CHAPTER II

FORCE

CHAPTER III

WORK AND ENERGY

CHAPTER IV

ROTATION—GRAVITATION

CHAPTER V

FRICTION—MACHINES

CHAPTER VI

MEASUREMENT OF LENGTH, MASS, AND TIME

CHAPTER VII

GENERAL PROPERTIES OF MATTER

CHAPTER VIII

PROPERTIES OF GASES

CHAPTER IX

PROPERTIES OF LIQUIDS

CHAPTER X

MOLECULAR PHENOMENA IN LIQUIDS

CHAPTER XI

PROPERTIES OF SOLIDS

BOOK II

HEAT

CHAPTER I

TEMPERATURE AND EXPANSION

CHAPTER II

CALORIMETRY

CHAPTER III

CHANGE OF STATE

CHAPTER IV

CONVECTION AND CONDUCTION

CHAPTER V

CONNECTION BETWEEN WORK AND HEAT

BOOK III

PERIODIC MOTION AND SOUND

PART I—PERIODIC MOTION. WAVE MOTION

CHAPTER I

VIBRATORY MOTION

CHAPTER II

WAVE-MOTION

PART II—SOUND

CHAPTER III

PRODUCTION AND PROPAGATION OF SOUND-WAVES

CHAPTER IV

VIBRATIONS OF STRINGS, RODS, AND COLUMNS OF GAS

CHAPTER V

RESONANCE AND MAINTENANCE OF VIBRATIONS

CHAPTER VI

AUDITION

BOOK IV

LIGHT

CHAPTER I

REFLECTION

CHAPTER II

REFRACTION

CHAPTER III

OPTICAL INSTRUMENTS

CHAPTER IV

DISPERSION

CHAPTER V

LIGHT CONSIDERED AS A WAVE-MOTION

CHAPTER VI

PHOTOMETRY

CHAPTER VII

RADIATION

CHAPTER VIII

POLARISATION

BOOK V

MAGNETISM AND ELECTRICITY

CHAPTER I

MAGNETS AND MAGNETIC FIELDS

CHAPTER II

TERRESTRIAL MAGNETISM

CHAPTER III

MAGNETIC INDUCTION

CHAPTER IV

ELECTROSTATICS

CHAPTER V

THE ELECTRIC CURRENT

CHAPTER VI

THE MAGNETIC FIELD PRODUCED BY CURRENTS

CHAPTER VII

INDUCED CURRENTS

CHAPTER VIII

ELECTROLYSIS

CHAPTER IX

PASSAGE OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY

CHAPTER X

ELECTRIC OSCILLATIONS AND WAVES

CHAPTER XI

THERMO-ELECTRICITY

BOOK I

MECHANICS AND PROPERTIES OF MATTER

PHYSICS

CHAPTER I

DISPLACEMENT

1. Position.—In order to specify the position of a point on a plane surface, such as a sheet of paper, we may give its perpendicular distance from two intersecting reference lines drawn on the surface. Although it is not essential, it is in general convenient to take the reference lines, or axes as they are called, at right angles, when we are said to use rectangular co-ordinates. The distances of the point from the axes are called the co-ordinates of the point, and when we know the position of the axes and the co-ordinates of a point we can immediately locate the point. Thus suppose we take the left-hand and bottom edges of a sheet of paper as the axes and are given that the distance of a point from the lower edge is 3 inches and the distance from the left-hand edge is 4·5 inches; then if we draw a line parallel to the lower edge and at a distance of 3 inches, it is obvious that the given point must lie somewhere in this line. In the same way if we draw a line parallel to the left-hand edge and at a distance of 4·5 inches from it, the point must lie on this line. Hence the point must be at the only point which is common to the two lines, that is at their intersection. Further, it is evident that there can be only one point which has the given co-ordinates.

Rectangular co-ordinates.

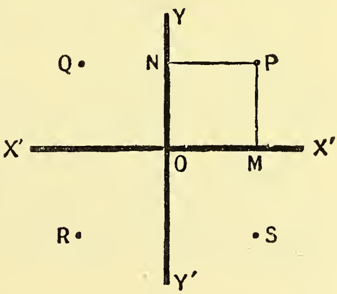


FIG. 1.

If XOX′ and YOY′ (Fig. 1) are the co ordinate axes, the co-ordinates of a point P are OM and ON, the co-ordinate OM being called the *abscissa* of the point P and ON the *ordinate*. A common convention is to use the letter x to indicate the abscissa of a point and the letter y to indicate the

ordinate. It is also usual to take the abscissæ as positive when measured to the right of the point O where the axes intersect, this point being called the origin. The ordinates are usually taken as positive when measured above the point O. Thus the co-ordinates of the point P are both positive. The abscissa of the point Q is negative while the ordinate is positive. In the case of the point R both co-ordinates are negative, while for the point S the abscissa is positive but the ordinate is negative. All points included in the quadrant to the right of OY and above OX have both co-ordinates positive, and similarly for the other quadrants.

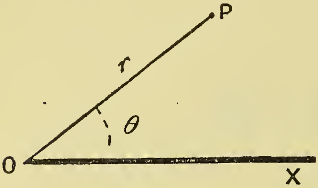


FIG. 2.

Another method of specifying the position of a point on a plane is sometimes convenient. Suppose we have a fixed origin O (Fig. 2) and a fixed line OX, then we can specify the position of the point P if we give the value of the angle XOP and the distance OP. In this case we are said to use *polar co-ordinates*. If we take OX as the one axis and a line through O perpendicular to OX as the other axis, we can change from polar co-ordinates (r, θ) to rectangular co-ordinates (x, y) by means of the following relations:—

Polar co-ordinates.

$$x = r\cos\theta; \qquad r = \sqrt{x^2+y^2}$$

$$y = r\sin\theta; \quad \tan\theta = \frac{y}{x}$$

2. Metric and British Units of Length, Volume, and Mass.—We have in the preceding section made no statement as to the units in which the co-ordinates of a point are expressed. Of course it is essential that the unit should either be explicitly stated or at any rate understood. In most scientific work the units of length employed are those of the metric system. The reason for this preference is that the subdivisions employed in this system are so much more convenient than those of the ordinary British system. The following table gives the relations between the more commonly employed units of length of the metric system, together with the values of the British units in terms of the metric units:—

Units of Length.

10 millimetres (mm.)	= 1 centimetre (cm.).
10 centimetres	= 1 decimetre (dm.).
10 decimetres	= 1 metre (m.).
1000 metres	= 1 kilometre (km.).

1 inch		= 2·540 centimetres.
1 foot	= 12 inches	= 30·479 centimetres.
1 yard	= 3 feet	= 0·9144 metres.
1 mile	= 1760 yards	= 1·6093 kilometres.

In the metric system the unit of volume is the volume of a cube each edge of which is of unit length.[1] Thus we have the cubic centimetre (c.c.), the cubic decimetre, which is called a litre, and the cubic metre. These units are related as follows to each other and to the British units:—

Units of Volume.

1000 cubic centimetres		= 1 litre.
1000 litres		= 1 cubic metre.
1 pint		= 0·5679 litres.
1 gallon	= 8 pints	= 4·5432 litres.
1 cubic foot	= 6·228 gallons	= 28·311 litres.

The units of quantity of matter or mass in the metric system are founded on the kilogram, which was originally intended to represent the mass of a cubic decimetre of water measured at the temperature of 4° C., this temperature being chosen since it is that at which water has its maximum density. The connection between the more generally employed units of mass are given in the following table:—

Units of Mass.

10 milligrams	= 1 centigram.
10 centigrams	= 1 gram.
1000 grams	= 1 kilogram.
1000 kilograms	= 1 metric ton.

1 ounce (avoirdupois)		= 28·3495 grams.
1 pound	= 16 ounces	= 0·45359 kilograms.
1 ton	= 2240 lbs.	= 1016·05 kilograms.

3. Speed, Velocity, Acceleration.—Suppose a small body to be moving in a straight line in such a way that the space passed over in each second is the same, then the body is said to be moving with a uniform *speed* or *velocity*. Since in such a case as that considered above the body not only passes over a given space in each second, but also it is moving in a given direction, to completely specify the motion we require to know the space passed over in unit time

[1] Originally the units of volume in the metric system were intended to be related to the units of length in this manner. Since, however, it is easier to measure the *weight* of some standard substance, such as water, than to determine the linear dimensions of a vessel, at the present day the litre is defined as the volume of a kilogram of water at a temperature of 4° C. For all practical purposes a litre as thus defined may be taken as equal to a cubic decimetre.

and also the *direction*[1] along which the motion is taking place. Occasionally the word speed is used only with reference to the space passed over in unit time, while the word velocity is used when in *addition* the idea of the direction of motion is considered. Thus two trains on different lines may be moving at the same speed, but if the lines are in different directions, their velocities will be different. A body which moves over, say, x centimetres in every second is said to have a speed of x centimetres per second, and this is sometimes written x cm./sec. or x cm. sec.$^{-1}$.

Unit of Speed.

If the body does not move in a straight line, the speed is the length of its *path* passed over in unit time. In such a case, although the speed may remain constant, since the path is continually changing its direction the velocity is not constant.

If the speed is not constant, so that equal distances are not passed over in successive equal intervals of time, the speed is said to be variable. In the case of variable speed, the speed at any given instant is obtained by considering the space passed over in such a small interval of time, which includes the given instant, that no appreciable change of speed takes place, and dividing the space passed over by the time.

In the case of a variable speed the amount by which the speed changes during a second is called the *acceleration*. If the speed is increasing the acceleration is said to be positive, while if the speed is decreasing the acceleration is negative. The acceleration may be uniform, in which case equal *changes* in speed take place in successive equal intervals of time, or it may be variable. When the acceleration is *uniform* and during a time t the speed changes from v_1 to v_2, the acceleration, a, is given by

Acceleration.

$$a = \frac{v_2 - v_1}{t} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

4. Velocity Curve. Space Traversed.—When the speed with which a body is moving is variable, a convenient method of representing this variation is to draw a curve or graph, such that for every point in the curve the abscissa represents a given instant of time and the ordinate represents the speed at that time. Since the use of such graphs will be of assistance when studying many problems, it is advisable to spend a little time in studying their properties. Although the results we obtain are quite general, it will facilitate matters if we refer to a particular graph, namely, that showing the relation between the speed

[1] Any quantity which for its complete specification requires that we know its direction, *i.e.* is a directed quantity, is called a *vector*. Those quantities, such as mass, which are completely defined if we know the amount only, are called *scalar* quantities.

of a body and the time, which is generally called the *velocity curve*, though the *speed* curve would be a more appropriate name.

If the body starts from *rest*, and as is usual we reckon the time from the instant that it starts, it is evident that the curve must go through the origin—that is, the point at which both the time and the speed are zero.

The speed curve for a body moving with constant speed is evidently a straight line parallel to the axis of time, for the ordinate of such a curve, which represents the speed, is always of the same length. Thus if OP (Fig. 3) represents a speed v, the straight line PP_1P_2 will be the velocity curve for a body moving with a constant speed v. Let the points T_1 and T_2 represent two times, so that T_1T_2 represents a given interval of time, and through T_1 and T_2 let us draw two perpendiculars T_1P_1 and T_2P_2 to meet the velocity curve. The area of the figure $T_1P_1P_2T_2$ is equal to $T_1T_2 \times T_1P_1$, but T_1T_2 represents the interval of time considered, and T_1P_1 represents the speed v. Hence the shaded area represents the product of the interval of time into the constant speed, but this product is the space passed over by the body. Hence the shaded area represents the space passed over. It is important to observe that we do not say that the area is *equal* to the space passed over, for it is obvious that an area cannot be equal to a distance, these two quantities being of a different nature. What we have shown is that there are as many units of area in the shaded figure as there are units of length passed over by the body in the time $T_2 - T_1$, a fact which may be indicated by saying that the area of the figure is *numerically* equal to the space passed over.

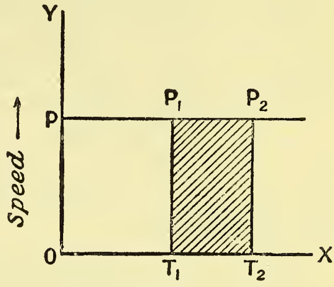


Fig. 3.

Space traversed represented by area.

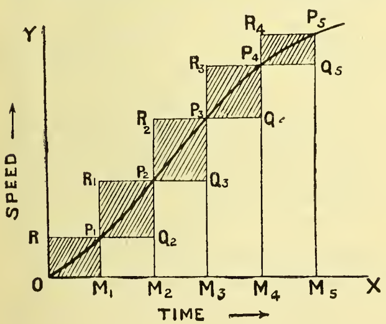


Fig. 4.

When the speed is not constant the velocity curve is no longer parallel to the time axis, for the ordinates will now be different at different times. Let $OP_1P_2P_3$, &c. (Fig. 4), be the velocity curve of a body starting from rest, and the

points M_1, M_2, M_3, &c., represent times of 1 second, 2 seconds, &c., after the start. Then M_1P_1 will represent the speed at the end of the first second of motion, M_2P_2 the speed at the end of the second second, and so on, while the speed varies continuously between each pair of ordinates. Let us now suppose that in place of the speed varying continuously, that the body moves with uniform speed during each second with the speed it actually possesses at the *commencement* of that second, but that the speed at the end of each second changes suddenly to the value it actually possesses at the commencement of the next second. The velocity curve would then be represented by the stepped line $OM_1P_1Q_2P_2Q_3P_3$, &c. Since the speed during each second is uniform, we can apply the result just obtained for uniform speed to each period of a second, and hence the space passed over in a given time will be represented by the area included between this stepped line, the axis of time and the two ordinates corresponding to the commencement and end of the interval.

Now for the particular curve shown in Fig. 4, where the speed increases with the time, the space traversed with the assumed motion will be less than the actual space, for the speed during each second is really *greater* than the speed at the commencement of the second. Next let us suppose that the body moves throughout each second at a uniform speed equal to the speed it actually possesses at the *end* of that second. The velocity curve will now be represented by the stepped line $RP_1R_1P_2R_2$, &c., and the space traversed will be represented by the area between this stepped line, the axis of time, and the ordinates as before, and this space will be greater than the space actually traversed. The difference between the spaces traversed in the two hypothetical cases is represented by the sum of the small rectangles shown shaded in the figure. Next suppose we take half-second intervals—that is, suppose that the speed remains constant for half a second and then changes abruptly. The difference in the space passed over in the two hypothetical cases is now shown by the shaded rectangles in Fig. 5. It will be observed that the area of these rectangles is only half the area of the corresponding rectangles when we took whole second periods. Hence we infer that by taking sufficiently small intervals the difference between the areas included by the two stepped curves

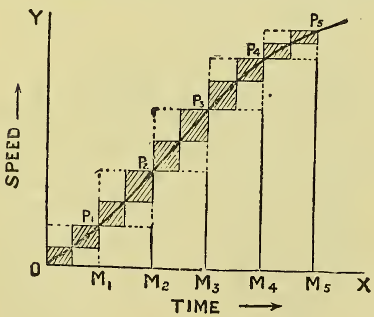


Fig. 5.

can be made as small as we please, while the true space traversed is always intermediate between those corresponding to the two stepped curves. Hence if the number of steps is very great both the stepped curves practically coincide with the actual velocity curve $OP_1P_2P_3$, &c., and the actual space traversed in any interval is represented numerically by the area included between the velocity curve, the axis of time, and the ordinates drawn through the points corresponding to the commencement and end of the interval considered.

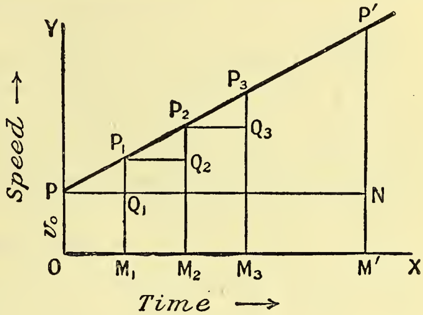


FIG. 6.

Let us now proceed to interpret what is the nature of the motion of a body for which the velocity curve is the straight line PP' (Fig. 6). In the first place the body at the instant when we start to reckon time is moving with a speed represented by OP, which we will call v_0. Further, the speed is obviously increasing, and if M_1, M_2, M_3 represent times of 1 second, 2 seconds, and 3 seconds from the start, the increase in the speed during the first second is P_1Q_1, the increase during the second second is P_2Q_2, and that during the third second is P_3Q_3. Now since $OM_1 = M_1M_2 = M_2M_3$, each representing 1 second, $PQ_1 = P_1Q_2 = P_2Q_3$. But the triangles PP_1Q_1, $P_1P_2Q_2$, $P_2P_3Q_3$ are similar. Therefore,

Velocity curve for uniform acceleration.

$$\frac{P_1Q_1}{PQ_1} = \frac{P_2Q_2}{P_1Q_2} = \frac{P_3Q_3}{P_2Q_3},$$

and hence $$P_1Q_1 = P_2Q_2 = P_3Q_3.$$

That is, the *increase* of speed during each second is the same, or in other words, the acceleration is constant. Hence we conclude that the velocity curve PP' represents the motion of a body which starts with a speed v_0 represented by OP, and moves with a uniform acceleration represented by P_1Q_1.

If M' represents a time t from the start, then NP' is equal to at, where a is the acceleration, for the speed will have increased by at in t seconds. The space traversed by the body in the time t is represented by the area of the figure $OPP'M'$, that is by the sum of the figures $OPNM'$ and $PP'N$. But the area of $OPNM'$ is equal to $OP \times OM'$, that is v_0t,

Space traversed when acceleration is uniform.

while the area of the triangle PP′N is $\frac{1}{2}$ NP′.PN, that is $\frac{1}{2}\,at \times t$. Hence the space s traversed in a time t is given by

$$s = v_0 t + \tfrac{1}{2}\,at^2 \quad . \quad . \quad . \quad . \quad (2)$$

The final speed at the time t is represented by P′M′, and is equal to $v_0 + at$. Hence calling this final speed v, we can write equation (2) in the form $s = v_0 t + \frac{1}{2}(v - v_0)t = \frac{1}{2}(v_0 + v)t$. Now $\frac{1}{2}(v_0 + v)$ is the *mean* speed during the interval t, and hence the space passed over is equal to the product of the time into the mean speed. If the body starts from rest $v_0 = 0$, and hence

$$s = \tfrac{1}{2}\,at^2 \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In this case $v = at$, and hence substituting v/a for t, we get

$$s = \frac{1}{2}a.\frac{v^2}{a^2} = \frac{v^2}{2a},$$

or

$$v^2 = 2as \quad . \quad . \quad . \quad . \quad (4)$$

An expression which gives us the speed acquired in a given *distance* by a body starting from rest and moving with a constant acceleration.

5. Composition and Resolution of Velocities and Accelerations.—The velocity of a body at any given instant can be represented by a straight line, for such a line may be drawn in the direction of the motion and of such a length that it contains as many units of length as there are units of speed in the velocity. The sense in which the movement takes place is usually indicated by an arrow-head placed on the line.

Velocity represented by a straight line.

Thus suppose that a ship is moving in a northerly direction with a speed of 10 feet per second, we can represent the velocity by a straight line ON (Fig. 7) 10 inches long, drawn upwards. This line will also represent the velocity of a man standing still on the ship's deck. If, however, the man walks across the deck towards the east at a speed of 4 feet per second, the velocity with which the man moves with reference to the earth is no longer represented by ON. From O draw OE in the easterly direction so that OE is 4 inches long, so that OE represents the velocity of the man with reference to the ship, and we have to consider what will be the motion of the man when the ship is in motion. It is evident that in a second the ship moves from O to N, and that when the ship is at N, the man will be at R, when NR is parallel and equal to OE. Further,

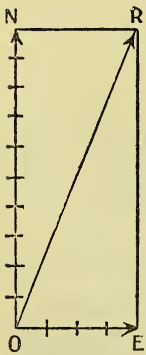


Fig. 7.

the actual path traversed by the man during the second is the diagonal OR, so that if in place of moving with a velocity represented by ON, and also a velocity represented by OE, the man had moved with a single velocity represented by OR, the result would be exactly the same. Hence the velocity OR is called the *resultant* velocity, the two velocities to which it is equivalent being called the *component* velocities.

Composition of velocities.

In the above we have taken the two component velocities in directions at right angles; the argument would, however, be exactly the same whatever the angle between the components, so that we have the following rule for finding the resultant of two velocities: From any given point O draw two straight lines to represent the given velocities both in magnitude and direction, and complete the parallelogram with these lines as adjacent sides, then the diagonal of the parallelogram drawn through the point O will represent the resultant velocity in magnitude and direction. When drawing lines to represent the component velocities, they must both be drawn away from O or both towards O. If they are drawn towards O, then the resultant will act along the diagonal in the sense towards O. The above construction for finding the resultant of two lines is known as the *parallelogram of velocities.*

In the parallelogram of velocities we are virtually finding the result of *adding* two velocities together. Now when we add two quantities such as volumes or masses together we have simply to add the numbers which represent the individual quantities; thus the sum of 10 lbs. and 4 lbs. is simply 14 lbs. The sum of a velocity of 10 ft./sec. north and a velocity of 4 ft./sec. east is not, however, a velocity of 14 ft./sec. The distinction between the two cases is that in the case of volume or mass there is no question of direction involved, so that we are dealing with scalar quantities, and hence the sum can be obtained by simple addition. A velocity on the other hand involves the question of direction, and is a *vector*. In the case of vectors, whatever their nature, the sum is always obtained by the parallelogram law as given above in the particular case of velocities.

Addition of vectors.

If we have two vectors of which the magnitudes are v_1 and v_2, and their directions make an angle θ, then since in Fig. 8 $\text{OC}^2 = \text{OA}^2 + \text{OB}^2 + 2\text{OA}.\text{OB}.\cos \text{AOB}$, we have, if the resultant vector is v,

$$v^2 = v_1{}^2 + v_2{}^2 + 2v_1v_2 \cos \theta \qquad . \quad . \quad . \quad (5)$$

If θ is zero, so that the vectors act in the same direction, we have $V = v_1 + v_2$; so that in this case the resultant is equal to the ordinary arithmetical sum of the components. The same result holds if $\theta = 180°$,

so that the vectors are in opposite directions, when the resultant is equal to the difference between the components. In the simple case where the components are at right angles, so that $\theta = 90°$, we have

$$V^2 = v_1^2 + v_2^2 \quad . \quad . \quad . \quad . \quad (6)$$

The parallelogram of velocities enables us to find the resultant of any number of velocities or other vectors, since by its means we can replace any two of them by their resultant. Next, this resultant can be combined with one of the remaining velocities, and so on till a single resultant is left.

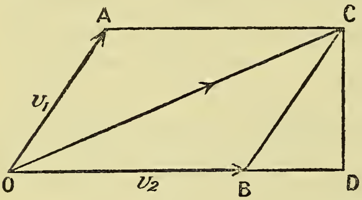


Fig. 8.

A consideration of Fig. 8 will show that the resultant velocity may be obtained without actually drawing the parallelogram OACB, for if from the point A at the extremity of the straight line OA, which represents in magnitude and direction one of the component velocities, we draw a straight line AC to represent in magnitude and direction the other component velocity, then the straight line OC joining the point O to the point C will represent in magnitude and direction the resultant velocity. Of course, the same result would be obtained by drawing through B, the extremity of the line representing the second velocity in the above case, a line BC to represent the first. This method of effecting the composition of velocities is generally referred to as the triangle of velocities, and leads to a convenient method of finding the resultant of a number of velocities, say, v_1, v_2, v_3, v_4.

Resultant of a number of velocities.

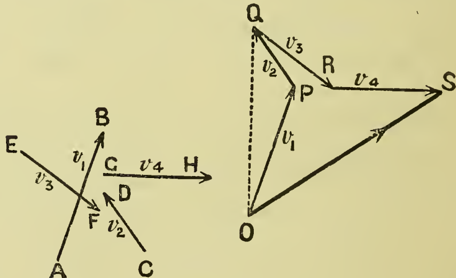


Fig. 9.

Suppose the given velocities are represented by the lines AB, CD, EF, GH (Fig. 9), then from any point O draw a straight line OP equal and parallel to AB, from P draw PQ equal and parallel to CD, from Q draw QR equal and parallel to EF, and from R draw RS equal and parallel to GH. Then the straight line OS will represent the resultant of the four velocities v_1, v_2, v_3, v_4. The velocity v_1 is represented by OP just as much as by AB; the same remark applies to the other velocities. Hence by the triangle of velocities the resultant of v_1 and v_2 is represented by OQ. Also the

resultant of OQ and v_3, *i.e.* QR, is represented by OR, and so on. It is important to note that the value of the resultant obtained is independent of the order in which we draw the lines representing the velocities. Thus in Fig. 10 the velocities have been combined in the order v_4, v_1, v_3, v_2; the resultant, however, is the same as before.

Resolution of velocities.

It is sometimes convenient to perform the inverse operation to the composition of velocities—that is, to replace a single velocity by two others which would have this single velocity as their resultant. In such a case we are said to *resolve* the given velocity into its components. The adjacent sides of *any* parallelogram having the given velocity for its diagonal will represent possible components, so that the resolution can be made in an infinite number of ways. In practice, however, it is nearly always convenient to resolve the velocity into a component in some given direction in which we wish to study the motion, and into a second component at right angles to this, which will obviously have no influence in the direction we are considering. Suppose, for instance, we require to find the time a sailing ship will take to go a certain distance in the teeth of the wind, being given that it sails at an angle of 45° to the direction of the wind with a speed of ten knots.[1] Let NO (Fig. 11) represent the direction of the wind, and OP the direction and magnitude of the velocity with which the boat sails. Resolve OP along ON, and at right angles to ON, *i.e.* along OE. Then we may consider that the boat moves with a velocity OM in the required direction, and with a velocity OL in a direction at right angles which has no effect on the space passed over in the required direction. As a matter of fact the boat would sail alternately in the direction OP and in the direction OQ. The resolved part of the velocity along ON would be the same in the two cases, but the resolved part at right angles to ON would be alternately in the direction OE, and in the direction OW; and hence the space

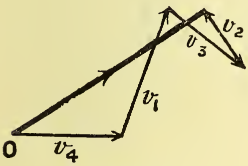


FIG. 10.

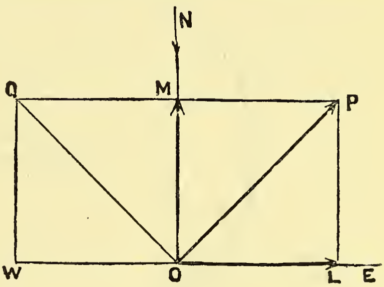


FIG. 11.

[1] The knot is the only special name we have for a unit of speed, and it represents a speed of one nautical mile per hour.

passed over at right angles to ON, due to each of these components, would be in opposite directions, and would neutralise each other.

Since an acceleration is a vector or directed quantity, we can compound and resolve accelerations in exactly the same way as velocities.

6. Motion in a Circle.—Suppose a particle is in motion along a circle of radius r, the speed being uniform and equal to v. Then when the particle is at P (Fig. 12) the direction of motion is along the tangent PP′, while when the particle gets to Q the motion is along the tangent QQ′. Hence, although the speed remains the same, the direction of motion has changed from PP′ to QQ′. From a point A draw AB parallel to PP′ and of such a length as to represent the speed at P, so that AB represents the velocity at P. In the same way draw AC to represent the velocity at Q. Since AB is parallel to

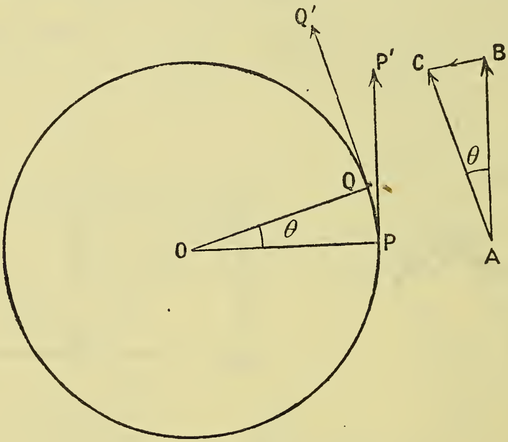


Fig. 12.

the tangent PP′ it is perpendicular to the radius OP. Similarly, AC is perpendicular to the radius OQ, so that the angle BAC is equal to the angle POQ or θ. If B C are joined, then from the triangle of velocities the velocity which compounded with AB will give AC is represented by BC, hence the change in velocity as the particle goes from P to Q is represented by BC. Now, since v is the speed of the particle along the circle, if dt is the very short time required to go from P to Q, we have

$$v = \frac{PQ}{dt} = \frac{r\theta}{dt}. \quad \text{Also } BC = AB.\theta = v\theta.$$

Hence combining these two results we get

$$BC = \frac{v^2 dt}{r} \text{ or } \frac{BC}{dt} = \frac{v^2}{r}.$$

Now BC represents the change in velocity during the time dt, and hence the acceleration a of the particle moving in the circle, that is the change in velocity by the time, is given by

$$a=\frac{v^2}{r} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

It is evident that this acceleration is constant in magnitude. The direction of the acceleration is along BC, which is parallel to a line bisecting the angle POQ. If P and Q are taken very close together, this bisector will coincide with the radius joining the particle to the centre. Hence we see that if a body moves round a circle of radius r with a uniform *speed* v, the *velocity* is not uniform, but there exists an acceleration of magnitude v^2/r, the direction of the acceleration being always along the radius joining the particle to the centre.

CHAPTER II

FORCE

7. Force. Newton's First Law of Motion.—It is a matter of common knowledge that a body which is at rest will remain at rest unless it is acted upon by some external agency. As, however, in many cases we are unable to perceive the exact manner in which the external agent acts, it is usual to employ the term force to denote any agency of this nature, and say that a body at rest remains at rest unless acted upon by some force due to some other body, such a force being called an external or impressed force. Newton extended the above idea, and showed that a body which is in motion will continue moving with uniform velocity unless it is acted upon by some *external force*. The direct experimental proof of this proposition is impossible, since we are unable to obtain a body which is entirely free from all external forces. What we do find, however, is that whenever a body does change its velocity in any manner whatsoever, then there is always an impressed force of some kind or other acting. Further, when problems in mechanics, in particular problems dealing with the motion of the planets, are solved on the assumption of the correctness of Newton's law, the results obtained are found to agree with observation. Thus we are justified in assuming the correctness of Newton's first law of motion, which is enunciated as follows :—

Every body continues in its state of rest or of uniform motion in a straight line, unless it be compelled by impressed force to change that state.

This law may also be considered as a definition of the term *force*.

8. Momentum. Newton's Second Law. Unit of Force.—The quantity of motion or *momentum* possessed by a moving body is proportional to the mass of the body and the velocity with which it is moving. Thus it requires a much more vigorous application of the brakes to stop a heavy motor-car than a light one, even if the speed with which the two cars are moving is the same. Again, it is more difficult to stop a car when it is moving quickly than when it is moving slowly, though a heavy train moving quite slowly is more difficult to stop than a motor-car moving quite quickly.

Momentum is product of mass into velocity.

Since the mass of a body cannot alter, the *change* of momentum of a body is due only to the change in velocity. Thus if a body of mass m moving in a straight line changes its speed from v_1 to v_2, the change of momentum is $mv_2 - mv_1$ or $m(v_2 - v_1)$. It must be remembered that the term velocity includes the idea of direction; thus a body which is moving with constant speed, but in a changing direction, as is the case when it moves in a circle, must be looked upon as having changing momentum.

Since change of momentum must be accompanied by change of velocity, and change of velocity is always due to the action of some impressed force, it follows that change in momentum must be due to impressed force, and we may use the change in the momentum produced by a given force as a measure of that force. This idea is embodied in Newton's second law of motion, which states:—

The change in momentum which takes place in unit time is proportional to the impressed force, and takes place in the direction of the straight line in which the force acts.

It must be noted, when considering the above law, that the total change in momentum produced by a given force depends on the *time* the force acts. Further, attention must be drawn to the fact that only the *change* in momentum is involved. Hence a given force will in a given time produce the same change in the momentum of a body (both of amount and in direction) whether this body is originally at rest or in motion in any direction at any speed. Again, if a number of forces act simultaneously on a body, each force will produce the same change in momentum of the body that it would produce if none of the other forces were in action.

If a force acts on a mass m for a time t, and the velocity of the body in the direction along which the force acts changes from v_1 to v_2, then the magnitude of the force is given by

$$F = \frac{m(v_2 - v_1)}{t}$$

But $(v_2 - v_1)/t$ is the acceleration (a) produced in the body. Hence

$$F = ma \qquad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus the unit force is such that it produces in unit mass unit acceleration. In the system of units founded on the centimetre as the unit of length, the gram as the unit of mass, and the second as the unit of time, which is called the C.G.S. system, the unit force is called a *dyne*, and is such that when it acts on a body having a mass of one gram it produces an acceleration of a centimetre per second per second in its own direction.

Unit force.

The force which when acting on a pound produces an acceleration of a foot per second per second is called a *poundal.*

The above units of force are generally called the *absolute units.* There is, however, another class of units of force called gravitational units, which are much more commonly employed by engineers. In this system the weight, or downward pull which the earth exerts in all bodies, is employed; the unit force being that with which the earth attracts a mass of a gram or a pound, as the case may be.

Gravitational units of force. Weight.

Experiment shows that all bodies at the surface of the earth, when allowed to fall freely, move with an acceleration of about 981 centimetres per second per second, or 32·2 feet per second per second. The value varies slightly at different parts of the earth's surface, but the variation is not sufficiently great to be of any consequence in most engineering problems. Since the attraction of the earth on a gram, or the weight of a gram, causes an acceleration of 981 cm./sec.^{2}, and a force of one dyne produces an acceleration of *one* cm./sec.^{2} when it acts on a gram, it follows that the *weight* of a gram is equivalent to 981 dynes. Similarly, the weight of a pound is equivalent to 32·2 poundals. Or generally, if g is the acceleration due to gravity, *i.e.* the acceleration with which a body falls, the unit of force on the gravitational system is g times the unit in the absolute system.

If a force of w grams weight (or pounds weight) acts on a mass of M grams (or pounds), the acceleration a cm./sec.^{2} (or ft./sec.^{2}) produced is given by the equation

$$wg = Ma.$$

For the force is wg in absolute units, and such a force will by equation (8) be equal to Ma.

The use of gravitational units is often convenient, especially when the forces with which we are dealing are due to the weights of bodies, say when calculating the strength required in the rafters to support a roof, but until considerable experience in working problems has been acquired, the student will find errors will be avoided if he reduces all forces with which he has to deal to absolute units by multiplying the value in gravitational units by g.

9. Newton's Third Law of Motion. Stress.—In the preceding sections we have spoken of an applied force acting on a body without in any way considering how this force is produced. In general the force is produced by the action of some other body or bodies, though it is not always clear by what mechanism such bodies are able to exert force on one another. Thus by attaching a string to a body we are able by pulling to exert a force on the body, and it is quite clear that

the force is due to the tension produced in the string. When, however, a magnet is brought near a piece of iron it exerts a pull on the iron, but we are at present entirely ignorant as to the mechanism by means of which such force is communicated. In every case, however, whether there is a material connection between the body (A) which exerts the force and the body (B) on which the force acts, or not, it is found that there is always a simultaneous force exerted on the body A by the body B—that is, when a body A exerts an *action* on a body B, B simultaneously exerts a *reaction* on A. These two forces, action and reaction, are in reality simply two different aspects of the same affair, just as buying and selling are different aspects of one and the same transaction.

Every force is one component of a stress.

When two bodies exert force on each other, the mutual action is called a *stress*, and the action and reaction considered above constitute a stress, and are related to one another in a manner which is expressed in Newton's third law of motion. This law says:—

To every action there is always an equal and contrary reaction; or, the mutual actions of any two bodies are always equal and oppositely directed.

It follows that all forces are of the nature of a stress between portions of matter, and that every force must of necessity be accompanied by an equal and oppositely directed force.

If you press your finger on the table, you feel the table pressing your finger. In the case of a horse towing a boat, the forward pull exerted by the horse on the tow-rope is exactly equal to the backward pull exerted by the tow-rope on the horse. Many people find a difficulty in accepting the above statement with reference to the equality of the action and reaction in the case of a horse towing a boat, since they think that if the force exerted by the horse on the rope were not a *little* greater than the backward force exerted by the rope on the horse, the boat would not progress. In this case we must, however, remember that, as far as their *relative positions* are concerned, the horse and the boat are *at rest*, and form a single body, and the *action* and *reaction* between them, due to the tension on the rope, must be equal and opposite, for otherwise there would be relative motion, one with respect to the other. The horse obtains the necessary purchase to move both itself and the boat where its feet touch the ground. At these points the horse's hoofs exert a force which has a component in a backward direction, the corresponding reaction of the ground having a component in the forward direction; and it is this component which produces the motion of the horse and boat.

10. Composition and Resolution of Forces.—In order to completely specify a force we must give (1) its point of application, (2) its direction, and (3) its magnitude. Since a straight line can be drawn from the point of application of the force in the direction of action, and having a length proportional to the magnitude of the force, it is very often convenient to use such a line to represent the force. The sense in which the force acts along the line is usually indicated by an arrow head.

Since a force is a vector quantity, we may add together or compound two forces which act at the *same point* by the parallelogram construction applied to the straight lines drawn to represent the forces. In the same way, if there are a number of forces acting at a point the resultant can be found by a construction similar to the polygon of velocities. In this case, however, it must be remembered that the side which closes the polygon only represents the resultant in magnitude and direction. The point of application of the resultant is of course the same as the point of application of the component forces.

Just as in the case of a velocity, a force can be resolved into components (see § 5).

The usual case is to resolve a force into components along two directions at right angles to one another. As an example of the resolution of forces, we may take the case of a boat sailing in any direction except directly before the wind. Let BR (Fig. 13) represent a boat, and AD the plan of the sail. If WAC is the direction of the wind, and we take AC to represent the force the wind would exert on the sail if it were placed at right angles to the direction of the wind, we must resolve this into components, one (AE) perpendicular to the sail, which is alone efficacious as far as the action of the wind on the sail is concerned, and the other parallel to the sail AD. The force AE has now to be resolved along and across the boat. The component AG is alone effective as far as the headway is concerned. The component AF, at right angles to the course of the boat, tends to make the boat travel through the water in a direction at right angles to its length, *i.e.* it produces leeway.

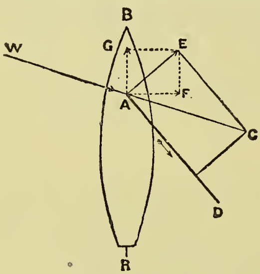


Fig. 13.

11. Moment of a Force. Torque.—If a force act on a body which can only rotate about a fixed axis, and the direction of the force passes through the axis, the force will not tend to rotate the body about that axis. If, however, the direction of the force does not pass through the axis, the force will tend to produce rotation, and the effect it will

produce is proportional to the magnitude of the force and to the perpendicular distance between the axis and the *direction* of the line of action of the force.

It will be noticed that it is the perpendicular distance of the *direction* of the line of action of a force from the axis of rotation, and not the distance from the point of application of the force to the axis, which settles the amount of turning power of a force. Thus let AB (Fig. 14) be a rigid body capable of rotating about an axis through A perpendicular to the paper. Then the turning power of a force acting along F_1 and applied at B is much less than that of an equal force acting along F_2, although the distance between the point of application and the axis is the same in each case. The turning power or torque depends on the magnitude of the force and the perpendicular distance between the axis and the direction of the force, *i.e.* on AC or AD in the above figure.

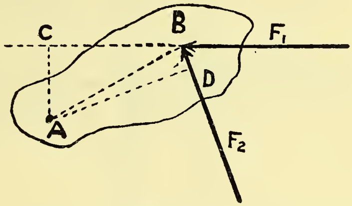


Fig. 14.

The product of the magnitude of a force into the perpendicular distance between the direction of the force and any given axis is called the *moment* of the force or *torque* about that axis. Hence in the above example the moments of the forces are $F_1 \cdot AC$ and $F_2 \cdot AD$ respectively. The moment of a force is positive if it tends to produce rotation in the positive (*i.e.* anti-clockwise) direction, and negative if it tends to produce rotation in the negative direction.

12. Composition of Parallel Forces. Couple.—When considering the composition of forces in § 10, we supposed that the component forces all acted at a single point; in other words, we considered the forces as acting on a particle. If, however, we are dealing with an extended body, the forces which act on it need not necessarily have a common point of application. The consideration of the general case in which we may have forces acting in different planes is beyond the scope of this book. We may, however, consider the case where we have two forces acting in the same plane. If the directions of the lines of action of the forces are not parallel they will intersect, when their resultant will have the same magnitude and direction as if the two forces acted on a particle placed at the point of intersection, which resultant can be found by the ordinary parallelogram, law. If the direction of the lines of action of the forces are parallel, then another method of finding the resultant has to be adopted.

First take the case when the two parallel forces, P and Q (Fig. 15), act in the same sense. The magnitude of the resultant R is obviously

P + Q, and the direction is parallel to the two forces, but we have to determine whereabouts with reference to the two forces it acts. Now if R is really the resultant of P and Q, its moment about any point must be equal to the sum of the moments of the two components, and we now proceed to try whether we can find a line of action for R which shall satisfy this condition. Take away point O and draw OACB perpendicular to the lines of action of the forces, and call the distances OA, OB, and OC, a, b, and c respectively. Then the moment of P about O is Pa, that of Q is Qb, and that of the resultant is Rc or $(P+Q)c$. Hence the condition that the moment of the resultant may be equal to the sum of the moments of the components is

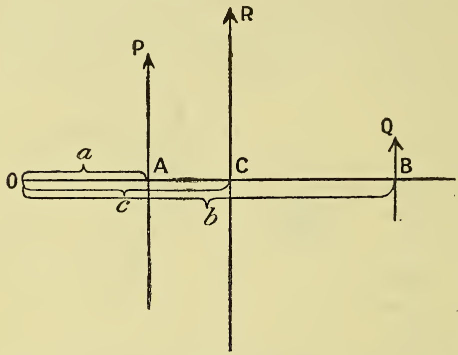


FIG. 15.

$$(P+Q)c = Pa + Qb$$

or

$$P(c-a) = Q(b-c)$$

$$\therefore \qquad \frac{b-c}{c-a} = \frac{P}{Q}.$$

But $b-c$ is equal to CB, and $c-a$ is equal to AC, and hence the condition is that $\frac{P}{Q} = \frac{CB}{AC}$, so that the resultant acts between the two parallel forces, the distance of its line of action from the components being inversely as these components. It will be observed that this result is quite independent of the position assumed for the point O about which we took moments.

Next suppose that the parallel forces are in opposite directions, and that the larger force P acts upwards and the smaller Q acts downwards. The magnitude of the resultant is P − Q, and to find its position we take moments about any point O as before, when we have

$$(P-Q)c = Pa - Qb$$

or

$$P(a-c) = Q(b-c)$$

$$\therefore \qquad \frac{b-c}{a-c} = \frac{P}{Q}.$$

Now no position for C can be found *between* A and B which shall satisfy this condition. If, however, the point C is taken on the side

of A remote from B, as shown in Fig. 16, then $b-c$ is equal to CB, and $a-c$ is equal to CA, and the condition can be satisfied, and is again independent of the position of O. Hence the resultant of two opposed parallel forces is equal to the difference of the forces, and acts on the side of the larger force remote from the smaller force, so that the distance of its line of action from the lines of action of the components is inversely as these forces.

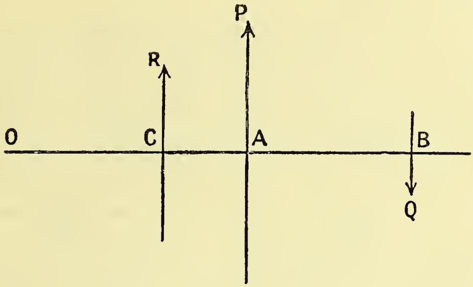


Fig. 16.

Resultant of two equal and opposite parallel forces.

There is a particular case of opposed parallel forces, namely, that in which the forces are of equal magnitude. In this case there is no *single* force which will be equivalent to the two forces. Two such equal and opposite parallel forces are called a *couple*, and while they have no tendency to move the body on which they act as a whole, they do tend to cause rotation, or in other words, they exert a *torque*.

The amount of torque exerted by a couple, or moment of the couple, is measured by the product of the magnitude of *one* of the forces into the perpendicular distance between their lines of action. It can easily be shown, by an argument similar to that given above when dealing with the resultant of unequal parallel forces, that the sum of the moments of the two forces which institute a couple about *any point* is equal to the moment of the couple.

13. Equilibrium of a Particle. Triangle and Polygon of Forces.—When the forces which act on a body are so balanced that they produce no acceleration either linear or angular in the body—that is, do not alter its state of motion, of translation or rotation—they are said to be in equilibrium. It is only when the forces which act on a body are in equilibrium that the body can *remain* at rest.

When we are dealing with the equilibrium of a *body* we have to consider both linear acceleration and angular acceleration (see § 19), for in general the motion of the body will consist both of a displacement as a whole and also a rotation. If, however, we confine our attention to a body in which we can only have displacement as a whole, and which is so small that all the forces in play must act at the same point, we are said to deal with a *particle*, and the conditions for equilibrium are more simple.

It is obvious that a particle acted upon by a single force cannot be in equilibrium.

For two forces acting on a particle to be in equilibrium, they must fulfil the following conditions: They must be (1) equal in magnitude, (2) act along the *same* straight line, (3) be of opposite sense. When referring to these conditions in future, we shall for shortness simply say that the forces must be equal and opposite, but it must be remembered that this is only an abbreviation for the above three conditions.

The condition that three forces acting on a particle may be in equilibrium is that any one of the forces must be equal and opposite to the resultant of the remaining two, for we may, if we please, replace any two of the forces by their resultant, when we should have reduced the problem to the equilibrium of two forces. The resultant of any two of the forces, say P and Q, must lie in the plane containing P and Q. Hence if there is to be equilibrium the third force, since it must be equal and opposite to this resultant, must also lie in the plane containing the other two

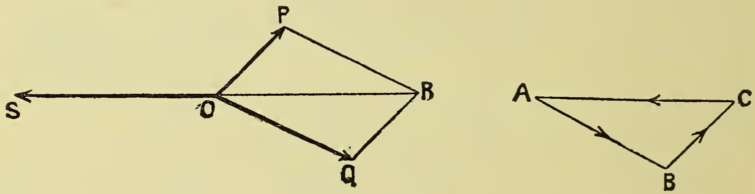


Fig. 17.

forces. Hence the first condition for equilibrium is that the three forces must all lie in one plane. As to the relations between the magnitude of the forces, the resultant of any two (P and Q) is represented by the diagonal OR (Fig. 17) of the parallelogram constructed on the lines OP and OQ as adjacent sides. Hence the other force must be represented in magnitude and direction by RO, or by OS, where OS is equal to OR and in the same straight line with it. Since QR is equal to OP, we may take QR to represent the force P in magnitude and direction. Then the three forces will be represented by OQ, QR, and RO, the sides of a triangle. Hence if it is possible to draw a triangle of which the sides are parallel or perpendicular[1] to the three forces and proportional to them in magnitude, the forces will be in equilibrium. It must be specially noticed that in drawing the triangle the sides must all be drawn in the same sense as the forces, so that when we place arrows on the sides to show in which sense the forces act, all the arrows may point the same way *round* the triangle, as shown at A B C in Fig. 17.

[1] Or more generally are equally inclined to the directions of the three forces.

The conditions of equilibrium for any number of forces acting on a particle are that the forces can be represented in magnitude and direction by the sides of a closed polygon taken in order, *i.e.* drawn in the same sense as the forces. This at once follows from the polygon of forces, for the resultant of all the forces but one is represented in magnitude and direction by the line joining the starting-point to the end of the last line drawn in the polygon, *i.e.* by the remaining side of the polygon, which by supposition represents in magnitude the only force not yet included, but is in an opposite sense.

14. Equilibrium of a Body.—In the case of a body the lines of action of the forces need not all pass through a single point, and in order that the body may be in equilibrium the forces must not tend to produce either translation or rotation. If the directions of all the forces pass through a single point they cannot produce rotation, and hence if they fulfil the conditions given in the preceding section for a particle they will be in equilibrium. If, however, the lines of action of the forces do not all pass through a point, then, in order that there may be no rotation, they must have no resultant moment tending to turn the body about any axis. The general condition for equilibrium is therefore that the sum of the moments of all the forces taken about *every* point must be zero, and that the forces can be represented in magnitude and direction by the sides of a closed polygon taken in order.

Since in most cases we shall only have to deal with forces acting in a plane, it is of interest to examine the condition for equilibrium in this case a little more fully. As by supposition the forces all act in a plane, it is evident that they can only tend to produce motion in this plane (by Newton's second law). Hence if we take two fixed lines not parallel (and preferably at right angles) in this plane, every possible translation must either be parallel to one or other of these lines, or else compounded of translations parallel to the two. Hence if the sum of the components of the forces when resolved parallel to these lines is zero, there will be no tendency to motion along either of these directions, so that there will be no translation. The condition for no rotation is that the sum of the moments about every point in the plane shall be zero. If both conditions are fulfilled there is equilibrium. If only the first condition holds, then there is rotation without translation, *i.e.* all the points of the body move in circles about a fixed point as centre; if the second condition alone is fulfilled, then there is translation without rotation, *i.e.* all the points of the body move with the same velocity in parallel paths.

CHAPTER III

WORK AND ENERGY

15. Work.—Suppose that when a force F acts on a body the point of application of the force moves through a distance s in the direction of the line of action of the force, the force is said to do work on the body, and the amount of work, W, performed is given by the product of the force into the displacement, or $W = Fs$.

If the body moves through a distance s in a direction opposed to the force F, then an amount of work Fs is said to be done *against* the force.

If the displacement does not take place in the direction of the line of action of the force, then the work is obtained by (1) resolving the displacement into a component along the line of action of the force, and another component at right angles to this direction, when the work is equal to the product of the force into the component displacement in its own direction, or (2) resolving the force into a component parallel to the displacement and a component at right angles to this displacement, when the work is equal to the product of the displacement into the component of the force in this direction. In the first case no work is done by the force as far as the component of the displacement perpendicular to its direction is concerned, for this displacement is at right angles to the force. In the second case there is no displacement in the direction of the component at right angles to the displacement, and hence this component does no work. It is easy by drawing a figure to show that the two methods of calculating the work given above lead to the same result.

Work, which is measured by the product of a force into a distance, must be clearly distinguished from torque or moment of a force, which is also measured by the product of a force into a distance. The distinction is that in the case of work the distance is measured *in the direction of the force*, while in the case of torque the distance is measured at *right angles to the direction of the force.*

In the C.G.S. system the unit of work is performed when a body acted upon by a force of an erg moves through a centimetre in the direction of the force. This unit is called an *erg.* The erg is a very small unit, so that to express amounts of work such as have to be dealt with in practice very large numbers

Unit of work.

would be required if this unit were used. Hence another unit called a *joule*, which is equal to ten million (10^7) ergs, is often employed.

In the British system of units the unit of work is done when the point of application of a force of a poundal moves through a foot in the direction of the force, and is called a *foot-poundal.*

The units of work in the gravitational systems (see § 8) are the work done in lifting a gram through a vertical distance of one centimetre, called a gram-centimetre; or more often the work done in lifting a kilogram through a vertical distance of a metre, called a kilogram-metre; and the work done when a pound is lifted through a vertical distance of a foot, called a foot-pound. In each case the lifting is supposed to be performed at the surface of the earth. Since the attraction or force exerted by gravity on a gram or pound, as the case may be, varies slightly at different parts of the earth's surface, these gravitational units also vary slightly, but the change is so small as to be of little consequence in all practical problems. It is evident that a gram-centimetre is about 981 ergs, and a foot-pound is about 32·2 foot-poundals, for the units of length on the absolute and gravitational systems are the same, and the weight of a gram is about 981 dynes, and the weight of a pound about 32·2 poundals.

The force with which the earth attracts a mass m is mg absolute units of force, and hence the work done when this mass is raised through a vertical height h is mgh absolute units of *work*. In gravitational units the work is simply mh, gram-centimetres, or foot-pounds, according as m and h are measured in grams or pounds and centimetres and feet as the case may be.

In § 4 we saw that if v is the speed of a body, the space described in a time t, when v is uniform, is given by vt, while if v is variable then we can obtain the space by drawing a curve giving the value of v at different times and measuring the area included between this curve, the axis of time, and two ordinates corresponding to the commencement and end of the interval considered. Now when a force F is uniform, the work done during a displacement s is Fs, and if F is variable we can, just as in § 4, show that if we draw a curve of which the abscissæ give the positions of the point of application of the force at different times, and the ordinates the corresponding values of F, then the area included between the curve, the x-axis, and the two ordinates drawn through points corresponding to any two positions of the point of application, will give the work done by the force during the displacement of its point of application between the two points chosen.

This method of measuring the work performed by a variable force was first introduced by Watt, who applied it to the steam-engine which he had invented, and since this is still the most important application of

the method, we may illustrate it by considering how the work done by the steam on the piston of a steam-engine can be calculated. The curve showing the connection between the force acting on the piston and the displacement of the piston is drawn by means of an instrument called an indicator, which is illustrated in Fig. 18. It consists of a cylinder A which is in communication with the engine cylinder through the tube G. A very light plunger B is a good but easy fit in the cylinder, and is connected by a rod C and a system of links with a lever EH. The upward movement of the plunger is resisted by a stout spring D. The pressure of the steam in the engine cylinder acting on the plunger forces it up against the spring, and the amount the plunger rises is a measure of the pressure in the cylinder. The motion of the plunger is magnified by the lever, and a pencil at E records the movement, and hence the pressure, on a drum F. This drum is given a reciprocating motion proportional to the movement of the piston of the engine, and hence the curve traced out by E gives the pressure in the cylinder for different positions of the piston. That is, the curve gives the force acting on the piston for different positions of the point of application, and hence allows of our calculating the work done on the piston by the steam.

Steam-engine indicator.

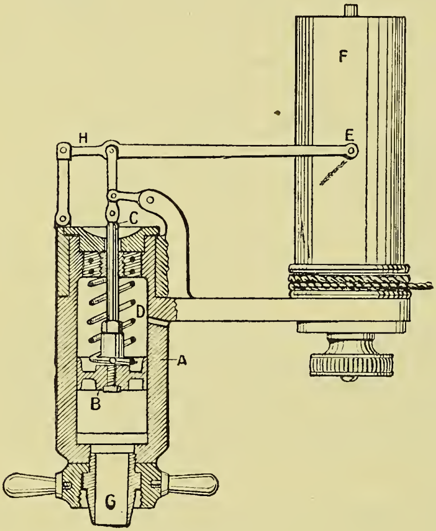


Fig. 18.

An example of the curve, called an indicator diagram, obtained from a steam-engine is given in Fig. 19. Starting from B when the piston is at the start of its stroke, the change in pressure during the outward stroke is indicated by the curve BCD, and hence the work done during this stroke is represented by the area of the figure OBCDM. During the return or exhaust stroke the piston is moving against the pressure, and the work done by the piston is represented by the area of the figure MDAO. The net work, therefore, done on the piston is represented by the area of the figure BCDEA. The above diagram represents the case of an engine which exhausts into the open air so that the pressure on the return stroke is not lower than the atmospheric pressure. If the engine

were to exhaust into a condenser, in which the pressure is less than atmospheric, the return curve would have the form shown dotted, and hence an additional amount of work represented by the area EDF would be done on the piston during each revolution of the engine.

16. Power. Activity.—When considering the subject of work it will be observed that nothing has been said as to the time occupied by the displacement of the point of application of the force, the amount of work performed being the same whether this displacement occupies a second, an hour, or a year. It is very evident, however, that there would be a very great practical difference between a steam crane which would lift a ton through a foot in a second, and another crane which took a minute about the same operation. Hence we have to introduce the idea of the *rate* at which an agent is capable of performing work. The rate of doing work, or the quotient of the work by the time taken to do

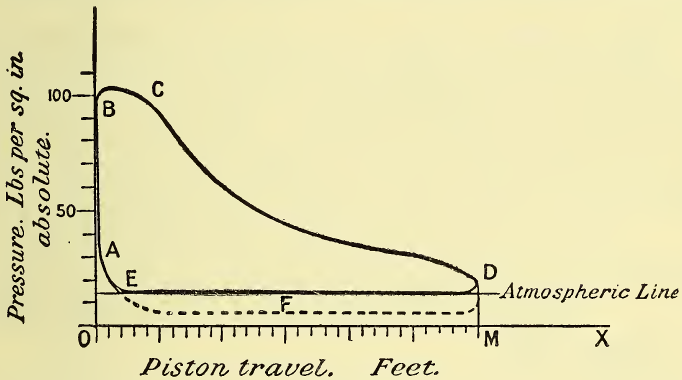


FIG. 19.

it, is called the *power* or *activity* of the agent which does the work. Thus in the example above, one crane has a power sixty times as great as the other.

An agent which does 550 foot-pounds of work in a second is said to exert a power of one *horse-power*, and this is the unit of power chiefly employed by engineers when dealing with steam-engines. Another unit of power is one joule per second, that is 10^7 ergs per second, and is called a *watt*. Since this unit is rather small, electrical engineers generally employ a unit which is equal to a thousand watts, and is called a *kilowatt*.

A horse-power is equal to 746 watts or ·746 kilowatt.

17. Energy.—When a body is so circumstanced that it is capable of doing work, it is said to possess *energy*. The amount of energy possessed by the body is measured by the work it can do while changing

to some standard state. Hence the units used to measure energy are the same as those of work, namely, the erg and foot-poundal in the absolute systems, and kilogram-metre and foot-pound in the gravitational systems.

Bodies which possess energy may be classified under two general heads: (*a*) those in which the energy is due to their movement, and hence as they do work they gradually lose their velocity, when the energy is said to be *kinetic*, and (*b*) those which possess energy due to their position or condition other than that of motion, when the energy is said to be *potential*.

Distinction between kinetic and potential energy.

A flying bullet has kinetic energy, and when it strikes a block of wood it does work in piercing the wood. A raised weight, such as a clock weight, on the other hand possesses potential energy, for as it descends it is capable of doing work. In the same way the wound-up spring of a watch possesses potential energy, for it can do work as it unwinds.

In the case of kinetic energy we can immediately calculate the energy of a body of mass m moving with a velocity v. Suppose the body brought to rest by an opposing force F, and that the space traversed by the body while this force is acting is s. The work done against the force is Fs, and hence the kinetic energy is Fs. Since the force is constant, the acceleration, a, it produces, which in this case is negative, must be uniform, and since in a distance s the speed decreases from v to o, the acceleration is $v^2/2s$ (p. 8). But $F=ma$, hence substituting for a we get $F=mv^2/2s$, or $Fs=mv^2/2$. Hence the kinetic energy of a body of mass m moving with a speed v is $\frac{1}{2}mv^2$.

Kinetic energy $=\frac{1}{2}mv^2$.

In the case of the potential energy of a body we cannot obtain a general expression such as that for the kinetic energy. In certain cases, however, the calculation is easy. Thus in the case of a weight of mass m, which has been raised through a height h, the work done in raising the weight, and which the weight can do while regaining its original position, is mgh, where g is the acceleration due to gravity. Thus the potential energy possessed by the raised weight with reference to its position before it was raised is mgh. The reason the calculation in this case is so simple is that the force mg acting on the weight, which is due to gravity, is independent of the height, at any rate for all practical purposes, and hence the work is simply the product of this force into the distance. In many cases, however, the force varies with the displacement of the body, and the calculation of the potential energy becomes more complicated. We may apply the method given in § 15 to calculate the potential energy of a stretched spiral spring, in which the force exerted by the spring is proportional to the extension. Let us measure the extension of the spring

Potential energy of raised weight.

along OX (Fig. 20), and the force exerted by the spring along OY. At O, where the extension is zero, the force is also zero, while when the extension is OM let the force be MP. Then since the force is by hypothesis proportional to the extension, the force for any extension is the ordinate of the straight line OP, so that when the extension is ON, the force is NQ. The work done when the spring is stretched by the amount OM is then equal to the area of the triangle OPM (§ 15). Hence the potential energy of the stretched spring is equal to half the product of the force exerted by the stretched spring into the extension.

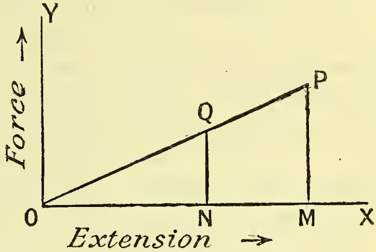


FIG. 20.

The energy of a body is capable of changing its form from potential to kinetic, and *vice versa*. Thus suppose a stone of mass m is supported on the edge of a cliff at a height h above the base of the cliff. The potential energy is equal to the work done in raising the stone through a vertical height h. The force with which the earth attracts the stone is equal to the product of the mass of the stone (m) into the acceleration which the force would produce (i.e. g). Hence in raising up the stone it has been moved through a space h against a force mg, and therefore the work done has been mgh, so that this is its potential energy.

Change of form of energy from kinetic to potential.

If now the stone be allowed to fall freely, it will gradually lose its potential energy, but will at the same time acquire velocity and hence kinetic energy. After it has fallen a distance s, its speed will be given by the equation $v^2 = 2gs$, and hence its kinetic energy ($\frac{1}{2}mv^2$) will be equal to mgs. The potential energy is now $mg(h-s)$, since the stone is now at a height $(h-s)$ above the ground. The sum of the kinetic and potential energies is therefore equal to $mgs + mg(h-s)$, which is equal to mgh, the original potential energy. When the stone reaches the ground its potential energy is zero, but the speed which it has acquired is now given by $v^2 = 2gh$, and hence the kinetic energy ($\frac{1}{2}mv^2$) is equal to mgh, the same value as the potential energy at the start.

Thus, although during the fall of the body there is a gradual change of potential energy into kinetic, the total energy remains constant.

If the stone were thrown vertically upwards with a speed v, then to start with the kinetic energy would be $\frac{1}{2}mv^2$. This would gradually diminish as the stone rose and lost speed; there would, however, be a gain of potential energy. When the stone is at the top of its flight it comes for an instant to rest, so that its kinetic energy is now zero. It

however possesses potential energy exactly equal in amount to its original kinetic energy.

18. Principle of the Conservation of Energy.—In the previous section we saw that although the *form* of the energy of the stone altered, so that it existed sometimes as potential energy, sometimes as kinetic energy, and sometimes as both together, yet the total quantity of energy was constant throughout. We shall in subsequent sections deal with many other forms in which energy can exist besides those already considered, and we shall find that these different forms of energy can be converted from one kind to the other.

When the stone reaches the ground it is brought to rest by impact with the ground, when it will apparently have lost both its kinetic energy and its potential energy. It is, however, found that the energy has not been *lost*, but has been transformed into another form, namely, that of heat, so that both the stone and the part of the earth it struck are warmer than they were before the impact. Joule has indeed shown, as we shall see later, that in every case a given amount of work entirely spent in producing heat always produces the same quantity of heat, no matter how the conversion takes place. For instance, a given number of ergs of work which can be obtained from stopping a moving bullet, and hence destroying its *kinetic* energy, will produce a certain quantity of heat. The same number of ergs of work done in rubbing a button on a piece of wood will produce exactly the same quantity of heat. The energy in the form of heat possessed by a body is supposed to be due to the motion of its particles. If this be so, then the kinetic energy of the stone moving as a whole is converted, by impact with the ground, into kinetic energy of the *particles* of the body and the earth near the point of impact, the particles moving (probably backwards and forwards with a vibratory motion) with, on the whole, a greater velocity than before.

These observations may be generalised, for in every case, without exception, it is found that the sum total of all the energy within any given boundary, through which energy is not allowed to pass, remains constant, although the energy within the boundary may be transformed into any of the many forms in which it is capable of existing.

Energy cannot be created or destroyed.

The above statement amounts to an enunciation of a doctrine which is practically the keystone of modern science, and is known as the *doctrine or principle of the conservation of energy.*

It follows that if the boundary considered includes the universe, the principle of the conservation of energy amounts to a statement that the sum total of the energy of the universe is a fixed unalterable quantity.

The principle of the conservation of energy also denies the possibility of "perpetual motion." By "perpetual motion" is meant the devising

of some arrangement so that energy in the form of mechanical work could be produced without energy in some other form being used up by the machine. Thus if an engine could be made to do work on external bodies for an indefinite time, and thus give out energy, without being supplied with energy from without, or diminishing the stock of energy in all its various forms which it originally possessed, we should have a means of creating energy, and this is in direct contradiction to the principle of the conservation of energy.

Although the total quantity of energy in the universe remains a constant quantity, so that whenever a given amount of energy in any one form disappears an exactly equal quantity of energy in some other form makes its appearance, yet there is, as far as man is concerned, a very important difference between the states of energy. Thus, to recur to the illustration of the stone on the edge of a cliff. By a suitable mechanical arrangement we can utilise some portion at any rate of the potential energy of the stone, and make the stone do work. For example, by attaching a string to the stone we can make it keep a clock going, and strike the hours on a bell. Again, we can let the stone fall, and utilise the kinetic energy it possesses by making it do work, say in driving in a pile. Suppose, however, it is simply allowed to fall on the ground. As a result of the impact with the ground its kinetic energy is destroyed, and the stone and the earth become warmer, and it is this heat which represents the lost kinetic energy; but in a very short time the heat of the stone, &c., will have diffused itself amongst surrounding objects. Energy in this form, that is, of uniformly diffused heat, is unavailable to man for the purpose of doing work; it is only when we have a body which is hotter than surrounding objects that we can utilise its heat energy to do work. Thus as far as we are concerned the energy of the stone has been wasted, for it is no longer available for doing work.

Availability of energy.

We shall find that in every transformation of energy there is some of the energy converted into heat, which becomes diffused throughout the universe, and is wasted. Thus, in the case of the raised stone being utilised to drive a clock, the friction at all the bearings causes some of the energy to be converted into heat. Heat also is produced each time the escape-wheel strikes the pallet of the detent, as well as the sound which we call the tick, and which we shall see is simply due to motion of the particles of the air, and this motion is gradually frittered away into heat by friction between the particles and other causes. Hence in this case all the energy of the raised stone eventually becomes changed into diffused heat, although in the meantime it goes through many transformations. Since in every transformation of energy from one form to another some of the energy becomes converted into uniformly diffused

heat, the total quantity of *available* energy of the universe is continually diminishing.

This continual degradation of energy, which accompanies every phenomenon with which we are acquainted, leads us to two conclusions: Firstly, since the quantity of unavailable energy is continually increasing, there must have been a time when none of the energy of the universe was unavailable, and before which no phenomenon, such as we are acquainted with, can have occurred, for every such phenomenon necessarily involves a degradation of energy. Secondly, there must necessarily arrive a time when all the energy will be unavailable, the whole universe having become a uniformly hot inert mass.

CHAPTER IV

ROTATION—GRAVITATION

19. Rotation. Angular Velocity. Angular Acceleration. Angular Momentum. Torque. Moment of Inertia.—When a body is rotating about an axis every particle of the body moves in a circle of which the plane is perpendicular to the axis, and the centre is situated in the axis. The angle swept out by the line joining any particle to the centre of its circular path is the same, and the number of radians[1] swept out in a second is called the angular velocity of the body.[2]

Angular velocity.

If the speed of rotation of a body is changing, then the change in angular velocity in a second is called the angular acceleration.

Angular velocity and angular acceleration are both vector quantities, for to *completely* define either of them we require not only to know the angle swept out in unit time, but also the *direction* of the axis about which rotation is taking place, and the sense of the rotation. The angular velocity of a body can conveniently be indicated by means of a straight line which is drawn in the direction of the axis of rotation, and has a length proportional to the magnitude of the angular velocity. The sense of the rotation is usually indicated by an arrow placed in such a way that looking along the line in the direction of the arrow the rotation takes place in a clockwise direction. Thus the direction along the line representing the angular velocity is related to the direction of rotation as is the direction of motion of an ordinary corkscrew, or right-handed screw, related to the direction of rotation.

Let us suppose that a body rotating about an axis with an angular velocity o is built up of a number of particles of masses m_1, m_2, m_3, &c., the distances of these particles from the axis of rotation being r_1, r_2, r_3, &c. Then the linear speed of the first particle is or_1 and its linear momentum is m_1or_1. This linear momentum multiplied by the radius of the circle in which the

Angular momentum.

[1] A radian is equal to 57°·296, and is the angle subtended at the centre by an arc of a circle of which the length is equal to the radius. The circumference of a circle of radius r is $2\pi r$, and hence 360° is equal to $2\pi r/r$ or 2π radians.

[2] The speed of rotation of a body is often measured in *revolutions per minute*. Since one revolution is equal to 2π radians, the angular velocity is equal to the number of revolutions multiplied by 2π and divided by 60.

particle is moving is called the *angular momentum* of the particle. Hence the angular momentum is $m_1 o r_1^2$. Similarly, the angular momentum of the second particle is $m_2 o r_2^2$, and so on for the others. The angular momentum of the whole body is the sum of the angular momenta of the component particles. Hence angular momentum of body $= o(m_1 r_1^2 + m_2 r_2^3 + m_3 + m_3 r_3^2 + \&c.) = Io$, where $I = (m_1 r_1^2 + m_2 r_2^2 + m_3 r_3^2 + \&c.)$, and is called the *moment of inertia* of the body.

Moment of inertia.

Thus just as the linear momentum of a body is equal to the product of the mass into the linear velocity, so the angular momentum is equal to the product of the moment of inertia into the angular velocity. It must be observed that while the mass of a body is a constant, the moment of inertia of a body is not a constant, but may have a different value for every axis about which rotation takes place.

We have seen that the force acting on a particle is equal to the rate of change of linear momentum produced. The quantity which is equal to the rate of change of the angular momentum of a rotating body is called a *torque* or *moment of a force*, and corresponding to Newton's second law we have, when dealing with rotations, the following law :—

The change in the angular momentum of a rotating body is proportional to the applied torque, and the direction and axis about which the angular momentum is changed are the direction and axis in which the torque acts.

Thus if $\dot{o}$ is the angular acceleration of a body and I the moment of inertia about a given axis, then the torque T acting about this axis is given by

$$T = I\dot{o} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

If a torque T acts through an angle of θ radians, the work done is $T\theta$.

The kinetic energy of a rotating body is $\frac{1}{2}Io^2$. The correctness of this expression can at once be deduced by considering the body as built up of separate particles, and obtaining the sum of their kinetic energies exactly in the manner adopted above when obtaining an expression for the angular momentum.

If M is the mass of a body of which the moment of inertia is I, and if the whole mass of the body were concentrated at a point at a distance R from the axis of rotation, the energy, when the angular velocity is o, would be $\frac{1}{2}Mo^2R^2$. If then R is so chosen that the energy is the same as the actual energy of the body when moving with the same angular velocity, *i.e.* so that

$$\tfrac{1}{2}Io^2 = \tfrac{1}{2}Mo^2R^2$$

or

$$I = MR^2,$$

the quantity R is called the *radius of gyration* of the body. The radius of gyration may therefore be defined as the distance from the axis of rotation at which the whole mass of the rotating body would have to be concentrated in order that the energy of rotation should be the same as it is actually.

Radius of gyration.

Being given the moment of inertia I of a body about an axis which passes through its centre of gravity, suppose we require to find the moment of inertia about a parallel axis at a distance x from that through the centre of gravity. We may look upon the energy possessed by the rotating body as consisting of two parts—(1) its energy of rotation about its centre of gravity, *i.e.* $\frac{1}{2}Io^2$, and (2) the energy of a mass M concentrated at the centre of gravity and rotating about the given axis, *i.e.* $\frac{1}{2}Mx^2o^2$. Hence if I' is the moment of inertia about the new axis, we have

$$\tfrac{1}{2}I'o^2 = \tfrac{1}{2}Io^2 + \tfrac{1}{2}Mx^2o^2$$

or

$$I' = I + Mx^2$$

This expression enables us to calculate the moment of inertia about any axis, that about a *parallel* axis through the centre of gravity being given.

The following table shows the relation between the different linear and angular quantities :—

Linear.		Angular.	
Distance . . .	s	Angle	θ
Velocity . . .	$v=\frac{s}{t}$	Angular velocity . .	$o=\frac{\theta}{t}$
Acceleration . . .	$a=\frac{v_2-v_1}{t}$	Angular acceleration .	$\dot{o}=\frac{o_2-o_1}{t}$
Mass	m	Moment of inertia . .	I
Momentum . . .	mv	Angular momentum .	Io
Force	ma	Torque	$T=I\dot{o}$
Work	Fs	Work	$T\theta$
Kinetic energy . .	$\frac{1}{2}mv^2$	Kinetic energy . .	$\frac{1}{2}Io^2$

20. Gravitation. — Newton first showed that the behaviour of bodies on the surface of the earth, as well as the movements of the planets round the sun, could be explained on the assumption that every portion of matter attracts every other portion of matter, and that the stress between them is proportional to the product of their masses divided by the square of their distance apart. Newton tested the truth of his law by showing that it correctly accounted for the force necessary to retain the moon in her orbit.

Thus if we assume that the orbit of the moon is a circle of radius R,

with the earth as centre, and the time the moon takes to complete the orbit is T, the linear speed with which the moon is moving is $2\pi R/T$, and the acceleration (§ 6) towards the centre of the orbit is $\left(\frac{2\pi R}{T}\right)^2 \div R$, or $\frac{4\pi^2 R}{T^2}$. Since the distance of the moon from the earth's centre is approximately 240,000 miles, and the time occupied in completing the orbit is 27·3 days, we have, reducing these quantities to feet and seconds respectively, that the acceleration is

$$\frac{4\pi^2 \times 240000 \times 5280}{(27{\cdot}3 \times 86400)^2} \text{ or } 0{\cdot}00899 \text{ ft./sec.}^2$$

Now the force acting on a given mass is proportional to the acceleration produced. Hence the forces exerted by the attraction of the earth on a given mass (a) when situated at the surface of the earth and (b) when at the distance of the moon are as the accelerations produced at these two distances from the centre of the earth.

Further, according to Newton's law of gravitation, the forces are inversely as the squares of the distances from the centre of the earth. Hence

$$\frac{\text{acceleration at distance of moon}}{\text{acceleration at surface of earth}} = \frac{r^2}{R^2}$$

where r is the radius of the earth and may be taken as 4000 miles. The acceleration at the surface of the earth is 32·2 ft./sec.^{2}, and hence the acceleration at a distance equal to that of the moon is

$$\frac{32{\cdot}2 \times (4000)^2}{(240000)^2} \text{ or } 0{\cdot}00894 \text{ ft./sec.},$$

which agrees with the value obtained as necessary to keep the moon in her orbit, the small difference being due to the approximate values we have assumed for the various quantities involved in the calculation.

The first direct proof of the attraction of two portions of matter of such a size that we can handle them was given by Cavendish. The difficulty of this experiment is very great owing to the fact that the mass of the largest body we can employ is excessively small compared to the mass of the earth, and hence the attraction between any two such bodies is very minute compared to their weight, that is, the force with which they are attracted by the earth.

The Cavendish experiment.

However, by using two large lead spheres to attract two small spheres which were carried at the end of a horizontal rod suspended by a very fine wire, Cavendish was able to produce a measurable twist of the lever owing to the attraction, and hence knowing the couple which was required to give unit twist to the suspension wire, he would calculate the magnitude of the attractive force.

Suppose a mass m when placed at a distance d from another mass m' attracts it with a force f, so that $f=\frac{mm'}{d^2}G$, where G is a constant which can be obtained from the above experiment. Now if R is the radius of the earth and M its mass, the force F with which the earth will attract the mass m at the surface of the earth is given by $F=\frac{mM}{R^2}G$. But the force with which the earth attracts m is mg, so that

$$mg=\frac{mM}{R^2}G$$

or

$$M=\frac{gR^2}{G}$$

Hence Cavendish's experiment, since it gives the value of the constant G, enables us to calculate the mass of the earth, and on this account Cavendish is often said to have weighed the earth.

Subsequent experiments made with more refined apparatus than that used by Cavendish have given $6{\cdot}66 \times 10^{-8}$ as the value of G when masses are expressed in grams and distances in centimetres.

21. Centre of Gravity. Mass Centre.—Suppose we have two particles at A and B (Fig. 21) of mass m_1 and m_2. Then they are each attracted towards the centre of the earth, and since the centre of the earth is at a very great distance compared with the distance AB, the forces exerted by gravity on the two particles will be sensibly parallel in direction. Hence we have the system consisting of the two particles acted upon by two parallel forces of magnitude m_1g and m_2g respectively. If we divide the line AB at C, so that

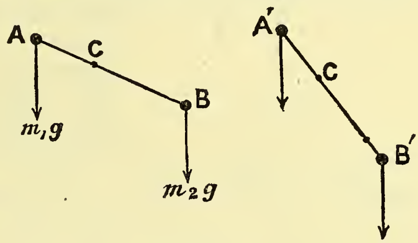


Fig. 21.

$$m_1g \times \text{AC} = m_2g \times \text{BC},$$

then the resultant of the two forces m_1g and m_2g will pass through C (see § 12). If the two particles are turned into any other position A′, B′, the distance AB between them being kept the same, the forces due to gravity will be inclined at a different angle to the line AB, but the resultant will still pass through the point C. Hence, whatever the position of the two particles A and B, so long as their distance apart remains the same, the resultant of the gravitational attraction of the earth on the two particles

always passes through a point C, which has a fixed position relatively to A and B.

If we have a system consisting of three particles, A, B, and D (Fig. 22) of mass m_1, m_2, and m_3 respectively, then the resultant of the earth's attraction on A and B acts at C, as in the previous case. We may now consider that we have *two* parallel forces, one of magnitude m_1g+m_2g acting at C, and the other of magnitude m_3g acting at D. Hence the resultant will pass through a point E, such that

$$(m_1g+m_2g)\text{CE}=m_3g\times\text{ED}.$$

Therefore the resultant of the earth's attraction on the three particles passes through E, and will still pass through E, however the three particles are turned, so long as their *relative* positions remain unaltered.

Proceeding in this way we might find, for a system consisting of any number of particles, a single point through which the resultant of all the forces exerted by gravity on the particles will pass, whatever the position of the system. As we may regard any solid body as built up of a number of particles, it therefore follows that it will be possible in the case of every solid body to find a point, and only one point, through which the resultant of all the forces exerted by gravity on the particles constituting the body must pass. This point is called the *centre of gravity* of the body. From this definition it follows that the weight of a body, which is simply the magnitude of the resultant of the forces exerted by gravity on the particles which constitute the body, always acts downwards in a vertical direction, passing through the centre of gravity.

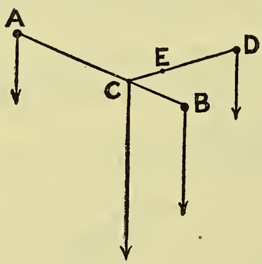


FIG. 22.

There cannot possibly be two centres of gravity belonging to one body, for if there were two, say G_1 and G_2, and the body was turned so that the line joining G_1 and G_2 was horizontal, then by definition the resultant of all the parallel forces due to gravity acting on the particles of the body passes through G_1; it also passes through a second point G_2, which is not on the line of the other resultant, which is impossible. Hence there can be only one centre of gravity.

The centre of gravity is a mathematical point, and it need not necessarily lie within the substance of the body. Thus the centre of gravity of a uniform ring lies outside the material of the ring.

That point related to a body or system of bodies which we have been considering above, namely, the centre of gravity, has properties of great importance quite independent of its connection with the gravitational

attraction of the earth, although for simplicity it is easiest to approach the subject from the gravitational point of view. Hence to indicate this fact what we have defined above as the centre of gravity is often called the *mass centre*. Although the consideration of the general properties of the mass centre would be beyond the scope of this book, we may examine one or two very simple cases to draw attention to the importance of this subject.

Mass centre.

Suppose we have two particles, A and B (Fig. 23), of mass m_1 and m_2 respectively, and that these are moving with uniform velocities v_1 and v_2, so that at the end of a second A comes to A′ and B to B′, and hence the line AA′ represents v_1 and the line BB′ represents v_2. Join AB and A′B′ and find the point C such that

$$\frac{AC}{CB}=\frac{m_1}{m_2}$$

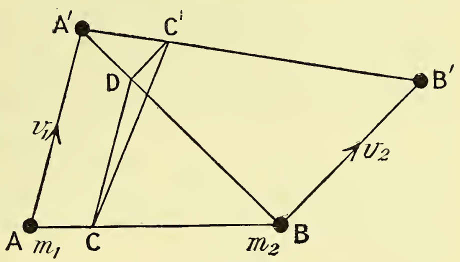


FIG. 23.

so that C is the mass centre of the bodies when they are at A and B. Join A′B, and through C draw CD parallel to AA′, and through D draw DC′ parallel to BB′, then C′ is the mass centre of the bodies when they are at A′B′.

For $$\frac{AC}{CB}=\frac{A'D}{DB}=\frac{A'C'}{C'B'}$$

Now $$\frac{CD}{AA'}=\frac{AC}{AB}=\frac{m_1}{m_1+m_2}$$

and AA′ is equal to v_1.

Hence $$CD(m_1+m_2)=m_1v_1$$

Similarly $$DC'(m_1+m_2)=m_2v_2$$

Now if we had a mass m_1+m_2 at the centre of mass C, and this mass moved with the centre of mass, its velocity would be represented by CC′. But the velocity CC′ is the resultant of the two velocities represented by CD and DC′. Therefore the momentum of the centre of mass, namely (m_1+m_2) CC′, is equal to the resultant of the momenta (m_1+m_2) CD and (m_1+m_2) DC′, that is to the resultant of the momenta m_1v_1 and m_2v_2. Hence the momentum of a mass m_1+m_2 moving with the velocity of the centre of mass would be equal to the sum[1] of the momenta of the two bodies which constitute the system. Proceeding in the same way we could extend the result to three bodies, and so on to the general case

[1] Since momentum is a directed or vector quantity, two momenta have to be added by the parallelogram law (§ 5), and their sum is what we have called their resultant.

where we had any number of bodies constituting the system, and it is obvious that we can, as far as the resulting momentum of the system is concerned, consider the whole system as concentrated at the centre of mass and that it moves with the velocity of the centre of mass.

Suppose we have a system, on the various component particles of which external forces act, then by considering the acceleration each force would produce in the particle on which it acts, we can show that the resulting acceleration of the centre of mass of the system will be the same as if all the forces acted upon a mass equal to the sum of all the masses concentrated at the centre of mass. It is not necessary to consider any forces which may be exerted between the individual particles which constitute the system. For suppose a particle A exerts a force F on another particle B in any direction, then by Newton's third law, particle B must exert a force F on particle A in the opposite direction. Hence, as far as their mutual action is concerned, the change in momentum of A is exactly equal and opposite to the change in momentum of B, and as far as the system which includes these two particles is concerned the momentum is unaltered.

The case of gravitation is only a particular case of this general proposition in which each of the forces is proportional to the mass on which it acts and the directions of all the forces are sensibly parallel.

Another important property of the centre of mass is that if a rigid body which is entirely unconstrained is struck a blow it will in general start moving in such a way that it has both translation and rotation. If, however, the direction of the force exerted by the blow passes through the centre of mass, then the motion set up will be one of translation only, there being no tendency to set up rotation about any axis whatever.

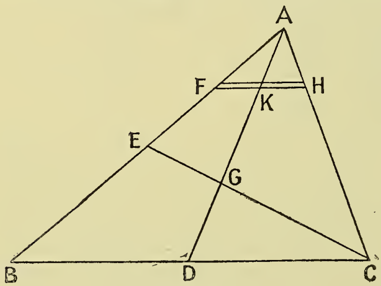


FIG. 24.

22. Calculation of the Position of the Centre of Gravity.

—The centre of gravity of a straight cylinder is obviously at the centre point of the axis. In the same way the centre of gravity of a thin uniform circular lamina is at the centre, and that of a square is at the intersection of the diagonals.

Centre of gravity of a triangle.

Let ABC (Fig. 24) be a triangular lamina, and suppose it divided into a large number of very narrow strips parallel to the side BC such as FH. Then the centre of gravity of each of these strips will be at its mid point, and all these mid points lie on the straight line joining A to the mid point of the side BC. Hence

the centre of gravity of the whole triangle must lie somewhere along AD. In the same way if E is the mid point of the side AB, the centre of gravity must lie somewhere on the line EC. Hence G must be the centre of gravity. That is, the centre of gravity of a triangle is at the intersection of the lines joining the angular points to the mid points of the opposite sides.

If in place of the triangle we had three equal masses m placed at the points A, B, and C, then the centre of gravity of B and C would be at D, and to find the centre of gravity of the three masses we may replace the masses at B and C by a mass of $2m$ at D, and then find the centre of gravity of this mass and the mass m at A. Thus the centre of gravity of the three masses must lie on the line AD. In the same way it is evident that the centre of gravity must lie on the line CE. Hence the centre of gravity of the three equal masses must lie at G, and must coincide with the centre of gravity of the triangle ABC.

Since G is the centre of gravity of a mass of $2m$ at D and a mass of m at A, we have $2m \times \text{DG} = m \times \text{AG}$, or $\text{AG} = 2\text{DG}$. That is the point G is at a *third* of the way up the line joining the mid point of a side to the opposite angle.

In an exactly similar way it can be proved that the centre of gravity of a pyramid on a triangular base is on the line joining the apex to the centre of gravity of the base, and at a *quarter* of the way up this line. It is then an easy extension to show that the same rule gives the centre of gravity of a pyramid on any other base, and hence of a cone.

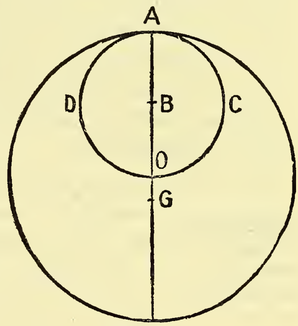


FIG. 25.

If we know the centre of gravities and masses of the parts into which a body may be divided, it is easy to calculate the centre of gravity of the whole body, or if we know the centre of gravity of a given body and the centre of gravity of a portion, we can at once calculate the position of the centre of gravity of the remainder. The principle on which such calculation is founded is that if we take *any* axis, the product of the mass of the whole body into the distance of the centre of gravity from this axis is equal to the sum of the products obtained by multiplying the mass of each part into the distance of its centre of gravity from the axis.

This result at once follows if we consider that the moment of the weight of the whole body about the axis must be equal to the sum of the moments of the weights of the parts about the same axis.

As an example of the application of this method, let us calculate the position of the centre of gravity of a circular lamina from which a hole has been punched as shown in Fig. 25.

If r is the radius of the disc, the radius of the hole ACD is $r/2$, and hence if m is the mass of the disc before the hole is punched, the mass of the portion removed is $m/4$. Hence if G is the centre of gravity of the portion left after the removal of the disc, we have, taking moments about A,

$$m \times AO = \frac{3m}{4} \times AG + \frac{m}{4} \times AB$$

or $$r = \frac{3AG}{4} + \frac{r}{8}$$

$\therefore$ $$AG = \tfrac{7}{6}r$$

We might have chosen an axis through the point O as that about which to take moments, when of course the moment of the whole disc would be zero, and hence

$$0 = \frac{3m}{4}OG - \frac{m}{4} \cdot OB$$

or $$OG = \frac{r}{6}$$

23. Equilibrium of a Body under the Action of Gravity.—Since the attraction exerted by the earth on a body is a vertical force passing through the centre of gravity, if the body is suspended by a string the direction of the string must pass through the centre of gravity. For the tension in the string must act along its length, and hence if the direction of the string did not pass through the centre of gravity, the tension and gravity would constitute a couple which would cause the body to rotate. In the same way, if the body is hung from a smooth peg, the centre of gravity must lie on the vertical drawn through the peg.

Experimental method of determining the centre of gravity.

The above considerations lead to an experimental method of determining the centre of gravity of a body. Thus suppose the body is suspended by a string, and the prolongation of the direction of the string is marked on the body, the centre of gravity must lie somewhere on this prolongation. If the suspending string is now attached to some other point of the body and the process repeated, a second line will be obtained which must also pass through the centre of gravity. Hence the centre of gravity must be at the *intersection* of the two lines.

Suppose that we have a body of the form shown in Fig. 26, which is suspended from a smooth peg A, then it is evident, as shown at (a) and (b), that there are two positions of the body in which the vertical through the centre of gravity G passes through the point of support. In (a) the centre of gravity is vertically *below* the point of support, while in (b) the

centre of gravity is vertically *above* the point of support. There is an important difference in these two cases, for if the body be slightly displaced, then in one case, as is shown at (*c*), the force of gravity tends to bring the body *back* to its original position, while in the other case (*d*) the force of gravity tends to *increase* the displacement. A body which is in the condition of equilibrium such as (*a*) where when slightly displaced it tends to return to its original position is said to be in *stable equilibrium*. While if when slightly displaced the tendency is for the displacement to increase, the body is said to be in *unstable equilibrium*. If a body is in such a condition that

States of equilibrium of a body.

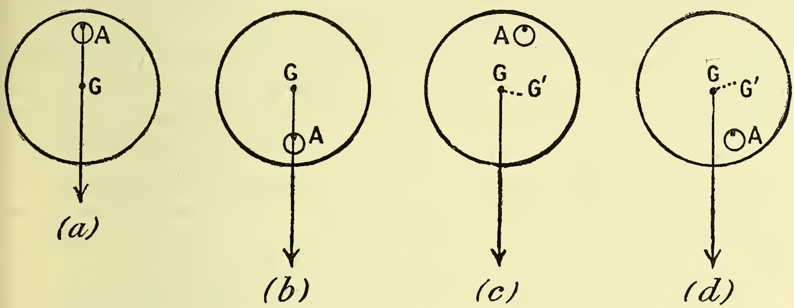


Fig. 26.

even when displaced the vertical through the centre of gravity passes through the point of support, then it is said to be in *neutral* equilibrium. A body supported at its centre of gravity or a sphere resting on a horizontal plane are examples of bodies in neutral equilibrium.

It will be noticed that when a body is in stable equilibrium, a displacement is accompanied by a rise of the centre of gravity (G′ to G in Fig. 26 (*c*)), hence there is an *increase* in the potential energy of the body. In the case of unstable equilibrium on the other hand, a displacement is accompanied by a fall in the centre of gravity and a *decrease* of the potential energy. In neutral equilibrium the potential energy is unaltered by a displacement.

CHAPTER V

FRICTION—MACHINES

24. Friction.—The idea of friction is familiar to every one, but in elementary books on mechanics it is usual to suppose that friction does not exist and then to calculate the motion of bodies and the working of certain simple machines on this assumption. Now friction can never in practice be eliminated, and in fact the results obtained on the supposition that friction is absent are often so obviously contrary to experience that unless special attention is drawn to the assumptions that have been made, the student is very apt to acquire a hearty contempt for the principles on which the calculations are made. For this reason we shall consider very briefly the subject of friction, and then draw attention, when considering the simple machines, to the effects of friction.

We first have to consider what is the nature of friction and how it can be measured. Suppose that a body C (Fig. 27), of mass m, rests upon a horizontal plane AB. Then, if no force except gravity acts, C will be in equilibrium under the action of two forces—(1) the weight mg of the body acting vertically downwards, and (2) the reaction of the plane, which must act vertically upwards and be equal to mg. Now, let a force P act on C, parallel to the surface AB. It is found that unless P exceeds a certain value the body still remains at rest. In these circumstances the body is in equilibrium under its weight mg, the force P and the reaction between its surface and the plane, which must now be inclined to the normal, and act in some such direction as CR′. This force along CR′ may be resolved into a reaction normal to the surface, *i.e.* along CR, and a force along CF which must, if there is equilibrium, be equal in magnitude to P. This force, which is brought into play when we attempt to slide one body over another, and which *always* acts so as to *resist* motion, is called the *friction* between the surfaces.

If the total *normal* pressure between C and the plane be Q, then it is found that C will commence to slide when the force P bears to Q a certain ratio, which is necessarily less than unity. This ratio is called the *coefficient of friction* between the body C and the plane AB. The value of the coefficient of friction is independent of the size of the surface of contact between C and AB, and of the pressure Q. It depends, however, on the nature

Coefficient of friction.

of the substances forming the two surfaces in contact, on the smoothness of these surfaces, and on the presence or absence of any lubricant, such as oil, fat, blacklead, &c., between the surfaces. The coefficient of friction has a value of about 0·5 for two pieces of wood when no lubricant is present, and of 0·1 for two metal surfaces which are greasy. In the case of well-lubricated metal surfaces the coefficient may be as low as 0·04.

If the force P (Fig. 27) is less than fQ, where f is the coefficient of friction between the surfaces, then there will be no motion, and the frictional resistance F will be equal and opposite to P. When P is just equal to fQ motion will be on the point of taking place, and the frictional resistance will have its maximum value (fQ). If P is greater than fQ motion will take place, but the moving force will be less than P, since, although when motion has commenced the frictional resistance is often a little less than fQ, yet friction still acts as a force tending to prevent

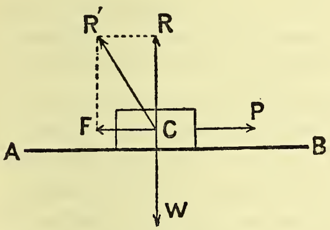


FIG. 27.

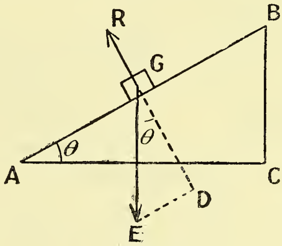


FIG. 28.

motion. Work has to be done against this frictional force when the body is displaced, and this work does not add either to the potential or kinetic energy of the body, but appears as heat which is developed at the rubbing surfaces.

If a body G (Fig. 28) of mass m is placed on an inclined plane AB, then, if there were no friction between G and the plane, the only forces acting would be the weight, which is a force of mg acting vertically downwards, and the reaction of the plane GR acting at right angles to AB. As these forces are not in the same straight line, the body would move down the incline. If, however, there is friction between G and the surface of the plane, the friction will tend to prevent motion, and till the plane has a certain slope the body will remain at rest. To find the maximum inclination (θ) of the plane to the horizontal we resolve the force mg into a component parallel to BA, which tends to produce motion, and a component normal to BA, which acts as the contact pressure. In the triangle DGE, the angle EGD is equal to θ, and ED is parallel to AB. Hence the component of mg parallel to BA is $mg \sin \theta$,

and the component perpendicular to BA is $mg \cos \theta$. If motion is just about to commence,

$$mg \sin \theta = fmg \cos \theta$$

$$\therefore \qquad f = \tan \theta \quad . \quad . \quad . \quad . \quad (10)$$

Hence if θ is less than the angle of which the tangent is equal to the coefficient of friction the body will remain at rest, while for larger values of θ the body will slide down the plane. This limiting value of the angle is called the *angle of repose*. If the coefficient of friction between one's boot and the road surface is ·6, the angle of repose is about 31°, and hence it would be impossible to walk on a slope of greater amount than this. If the coefficient of friction were 0·2, a value such as would be obtained between leather and greased wood, the limiting angle would be about 11°, and it would be impossible to walk up a greased plank at this or any greater angle.

Angle of repose.

When a body, such as a cylinder, rolls upon a plane surface there is no rubbing between the surfaces in contact such as occurs when one surface is dragged over another. There is, however, found to be produced at the place of contact of the cylinder and the plane a resistance which opposes the motion, and which is called rolling friction. The amount of the resistance in the case of rolling friction is in general very much less[1] than in the case of sliding friction for the same pressure between the two bodies. The resistance experienced in the case of rolling is due to the deformation of the cylinder and plane where they are in contact. The cylinder becomes slightly flattened and the plane becomes deformed, a small ridge being rolled up in front of the moving cylinder.

Rolling friction.

If F is the frictional force when a body is moved, the work which is done against friction when the body moves through a distance s is Fs. If the weight of the body is W and the coefficient of friction is f, the work done is Wfs. Thus the power expended in overcoming friction when the body is moved with a speed v is Wfv. Next let us suppose that the body is carried by a wheel and that the radius of the wheel is R and that of the axle is r. While the body moves through a distance s, the surfaces at the axle move through a distance sr/R. The weight supported by these surfaces is W, and if the coefficient of friction is f' the power expended against the friction at the axle is $Wf'vr/R$, and the ratio of this power to that expended in friction when sliding takes

Work performed against friction.

[1] Coulomb found that in the case of a cylinder of lignum-vitæ, 16 centimetres in diameter, when loaded with 1000 lbs. the resistance to rolling was only 3 per cent. of the resistance to sliding.

place is $f'r/fR$. Now since the axle can be lubricated and the surfaces are smooth, f' is generally very much smaller than f. Also r is much smaller than R. Hence the power expended against friction at the axle is considerably less than that expended if the body is simply dragged along the surface of the road. Of course the rolling friction between the wheel and the road has to be added, but even then there is a great advantage in favour of the wheel.

25. Friction-Dynamometer.—One of the applications of friction is to employ it to measure the power, or rate of doing work, of a machine, such as a steam-engine. A form of friction-dynamometer for this purpose is shown in Fig. 29. A pulley A with a flat edge is fixed to the shaft of the engine, and a strap BCD, on the inside of which blocks of wood are usually fixed, rests on the edge of this pulley. One end of the strap is attached to a spring balance E, by means of which the tension acting on this end of the strap can be measured, while a tension P, caused by a weight suspended on the other end, serves to keep the strap tight. The engine having been started, P is increased till the engine is exerting its maximum power; the work being done against the friction of the wooden blocks on the edge of the pulley. If r is the radius of the pulley and R the radius of the portion of a circle along which the strap lies, n the number of turns the pulley makes per second, and the reading on the spring-balance is W, then the sum of the moments of P and W about the axle is $(P-W)R$, and this must be equal to the moment of the friction about the same axis, since the strap and blocks are in equilibrium. Hence, if F is the frictional resistance,

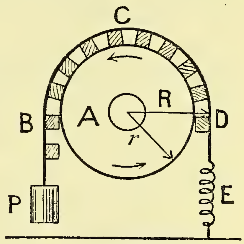


Fig. 29.

$$Fr = (P-W)R \quad . \quad . \quad . \quad (a).$$

Now the frictional resistance F acts tangentially to the pulley, and tends to check the motion. The distance through which the edge of the pulley moves against F during one second is $2\pi rn$. Hence the work done against friction in one second is

$$2\pi rn.F \quad . \quad . \quad . \quad . \quad (b),$$

and this is the power expended in friction. Substituting the value of F obtained from equation (a) in (b), we get that the power spent against friction is

$$2\pi rn.\frac{(P-W)R}{r}$$

or

$$2\pi nR(P-W).$$

If then the whole available power of the engine is spent in overcoming the friction of the dynamometer, and H is the number of units

of work per second, in the system in which R, P, and W are measured, which are equal to a horse-power, the horse-power will be

$$\text{HP} = \frac{2\pi nR(P-W)}{H} \quad . \quad . \quad . \quad . \quad (11)$$

in which n is obtained by counting, R by adding the thickness of the wooden blocks to the radius of the pulley, and P and W are obtained from the weights placed in the scale-pan and the reading of the spring balance respectively.

26. Simple Machines. Efficiency of a Machine. — A contrivance by means of which a force applied at one point gives rise to a force, often different in magnitude and direction to the impressed force, at some other point, is called a machine. A steam-engine, a hydraulic press, or a testing-machine, are all examples of machines; they however consist of a number of separate parts, each of which satisfies the definition of a machine given above, and which may be classified under various heads. Each of these separate parts is called a *simple machine.*

In studying these simple machines, we shall suppose that the machine is in equilibrium, so that the force impressed at one point is just balanced by the force impressed at some other point. One of these forces, which is impressed on the machine by some other body, is often called the *Power*, while the other is called the *Weight.* It must be carefully borne in mind that the term "power" in this relation has no connection with the same word used in § 16 to denote the rate of doing work, nor does the term "weight" necessarily mean that the machine is used to raise or support a mass against the attraction of gravity. Hence to save confusion it is better to call P the *effort* and W the *load.*

In finding the conditions which have to be fulfilled, in order that the simple machines may be in equilibrium, suppose at first that they work without friction. Under these conditions, if the machine receives a small displacement of such a nature that the connection between its parts is not in any way disturbed, the work done by the "effort" must, according to the principle of the conservation of energy, be equal to the work done on the "load." Hence, if P and W are the two forces which balance one another on a machine, and when the machine receives a small displacement the displacement of the point of application of P parallel to its line of action be p, and that of W parallel to its line of action be w, then the work done by P is Pp, and the work done on W is Ww.

Hence $$Pp = Ww,$$

or $$\frac{P}{W} = \frac{w}{p} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

From this we see that the displacements are *inversely* as the forces, so that if a small effort is to balance a large load, the distance

through which the effort acts must be large compared to the displacement of the load. This is often expressed by saying that "what is gained in power is lost in speed."

Velocity ratio of a machine.

The ratio p/w is often called the *mechanical advantage or velocity ratio* of the machine.

When we take into account the effects of friction, the relation between P and W is in general very much modified, but always in such a way that the value of the effort necessary to produce motion is greater than that given above. On the other hand, the machine will remain in equilibrium for all values of the effort, the load being given, between certain limits, these limits including the value calculated on the supposition that no friction exists.

If, owing to friction, a force $P+x$ is required to cause a displacement, then the work done by the effort is $(P+x)p$, while the work done on the load is Ww. The ratio $\frac{Ww}{(P+x)p}$, *i.e.* the ratio of the work performed by the machine to the work which has to be supplied to the machine, is called the *efficiency* of the machine. If the velocity ratio of the machine, i.e. p/w, is r, the efficiency may be written $W/(P+x)r$. If there is no friction, $x=0$ and the efficiency is 1. Whenever there is friction the efficiency must be less than unity, and the greater the friction the less the efficiency.

Efficiency of a machine.

If there were no friction, the load corresponding to $P+x$ would be $r(P+x)$. Hence the effect of friction is to reduce the load by an amount $(P+x)r-W$. Calling this reduction in the load F, it is found that in most cases $F=aW+c$, where a and c are constants for the given machine. The constant c represents the effort required to move the machine when there is no load.

27. The Lever.—A lever is a rigid bar, either straight or curved, which is capable of a motion of rotation about a fixed axis, called the fulcrum.

Since the lever, when in equilibrium, is under the action of three forces—the effort, the load, and the reaction of the fulcrum—it follows (§ 14) that the lines of action of all these forces must lie in one plane, and either be parallel or meet at a point.

The most direct way of obtaining the relation between the effort and load in the lever is to take moments round the fulcrum. If the lever is to be in equilibrium, these moments must be equal and opposite. Hence if a is the perpendicular distance between the fulcrum and the line of action of the effort, and b that between the fulcrum and the line of action and the load, we have

$$Pa=Wb,$$

or

$$\frac{P}{W}=\frac{b}{a} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

In the case where the lines of action of the forces are at right angles to the lines joining the points of application to the fulcrum, a and b represent the distances of the points of applications of the forces from the fulcrum, and are called the *arms* of the lever.

It is usual to divide levers into three classes according to the relative positions of the points of application of the forces and the fulcrum. In the first class the fulcrum is between the points of application of the forces. In the second and third classes the points of application of the forces are both on the same side of the fulcrum, and hence they must act in opposite directions. If the load, that is the force which the lever is employed to exert, is next the fulcrum, the lever is of the second class, while if the effort is applied next the fulcrum, the lever is of the third class. In the first class the mechanical advantage may be greater, equal to, or less than unity. In the second class the mechanical advantage must be greater than unity, and in the third class it must be less than unity.

If, as is the case in the balance (§ 34), the fulcrum consists of a hard knife-edge resting on a hard plane, the effect of friction is excessively small. If the fulcrum consists of a pin working in a hole in the lever, as in the common pump handle, the friction between the lever and the pin acts tangentially to the surface of the pin. If we are dealing with a lever of the first kind, the pressure on the pin is $P+W$, and hence the friction is $(P+W)f$, where f is the coefficient of friction. If R is the radius of the pin, the moment of the frictional resistance will be $(P+W)fR$, and hence when the lever is just on the point of moving we have, taking moment about the centre of the pin,

$$Pa = Wb + (P+W)fR,$$

so that P has to be greater than when no friction is present. If $a=60$ inches, $b=10$ inches, $R=\cdot 5$ inch, $f=\cdot 1$, and $W=50$ lbs., the above formula gives 8·38 lbs. as the value of P. If there were no friction P would be 8·33 lbs., so that even in this case the effect of friction is quite small.

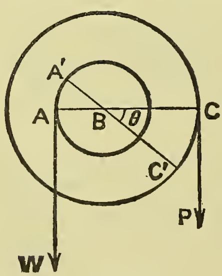


Fig. 30.

28. The Wheel and Axle. — This machine consists of two drums or wheels of different diameters fixed to the same axle. A rope coiled round the drum of smaller diameter carries the load, while the effort is provided by another rope coiled round the other drum in an opposite direction. It will be seen from Fig. 30 that the arrangement is virtually a lever with the fulcrum at the axis, B, about which the drums can turn. Hence $P \times BC = W \times AB$.

The advantage of this arrangement over the ordinary lever lies in the fact that while in the case of the lever the maximum distance through which the point of application of the load can be moved is limited to something less than the distance between this point of application and the fulcrum, no such limitation is involved in this case, the distance through which W can be moved simply depending on the amount of rope coiled on the drums.

In addition to the friction at the axle, there is a considerable amount of resistance due to the stiffness of the rope, so that additional work has to be done by the effort in winding the rope round one drum and unwinding it from the other.

The principle of the wheel and axle is used in the capstan and in the windlass. In these arrangements the effort, instead of being applied to the wheel by means of a rope wound round the circumference, is applied to the end of one or more rods which virtually form spokes of the wheel, the direction of the line of action of the effort being continually altered, as the machine turns round, so as to be always at right angles to the spoke.

29. The Pulley.—A pulley consists of a disc or wheel, called the sheave, mounted on an axle which is fixed to a framework called a block. The edge of the disc is usually grooved so that a cord can lie round it. If the block is fixed, then the direction of a force, but not its magnitude, may be changed by means of a pulley. If a tension P be applied to one end of a string which passes over such a pulley, then, since if we neglect the friction of the pulley and the stiffness of the string the tension is the same throughout the string, in order to keep the pulley in equilibrium the other end of the string must be pulled with a force P. This is at once evident, for if A and C (Fig. 31) are the points where the string leaves the pulley, and B is the centre of the sheave, then taking moments about B,

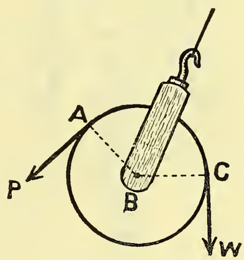


Fig. 31.

$$P \times AB = W \times BC.$$

But $AB = BC$. Hence $P = W$.

If the block of the pulley, instead of being fixed, is attached to the load, while one end (Fig. 32) of the string is attached to a fixed support, while the effort acts at the other end, P may be less than W. If, as is usually the case, the two portions of the string QA and CP are parallel, and the pulley moves through a distance h from the position ABC to the position A′B′C′, then the end of the string where P is attached will move

up through a distance $2h$, for the portion QA of the string has been shortened by a length h, and the point C has also risen through a height h. Therefore, while W has been raised through a height h, P must have moved through a distance $2h$, so that, *neglecting friction*, equating the work done in the two cases,

$$hW = 2hP$$

or

$$W = 2P.$$

In this expression, since the pulley itself has to be raised, we must include its weight in W.

There are several arrangements, called tackles, in which more than one pulley is used, but we shall only describe one of these, which is the

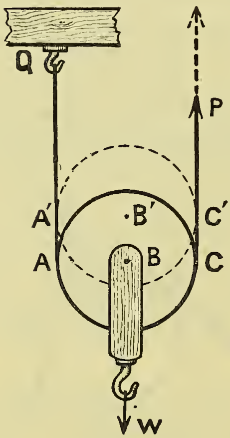


FIG. 32.

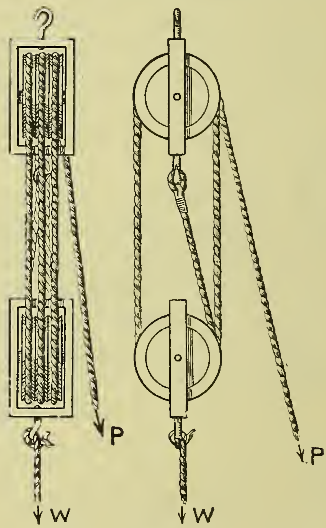


FIG. 33.

one most often employed in practice. It consists of two blocks, each fitted with several sheaves, which usually all turn on the same axle. One of the blocks is attached to a fixed point, while the other is attached to the "weight." One end of the string is attached to one of the blocks, it then passes round one of the sheaves of the other block, then over one in the first block, and so on till it has passed over all the sheaves. If the string passes n times from one block to the other, then we shall have

$$W = nP,$$

where W includes the weight of the movable block.

It will be seen in Fig. 33, where there are three sheaves in the movable block and the string passes six times from one block to the other,

that if the movable block, and therefore also W, is raised through a height h, then the free end of the string will have to move through a distance $6h$. Hence, *if we neglect friction*, the work done on W is hW, and that done by P is $6hP$. Therefore

$$hW = 6hP$$

or

$$W = 6P.$$

In general, if the cord passes n times between the two blocks, $W = nP$. Owing to friction of the sheaves on the axles, and the stiffness of the rope, the effort required to raise a given load is considerably greater than that calculated above. Thus to raise a ton with a tackle such as that shown in Fig. 32 would require an effort of about 5·5 cwt., so that the efficiency of the machine is $\frac{20}{5{\cdot}5 \times 6}$, or ·65.

30. The Differential Pulley or Barton Tackle.—A particular combination of pulleys which is much used in engineer's shops is illustrated in Fig. 34. It consists of two sheaves which are fastened together, and are of unequal diameter, in a fixed block A, and a single sheaved block B attached to the load. The edges of the sheaves are indented so as to prevent the chain, which is used in place of a rope, slipping. The chain is endless, and is wound round the sheaves in the manner shown.

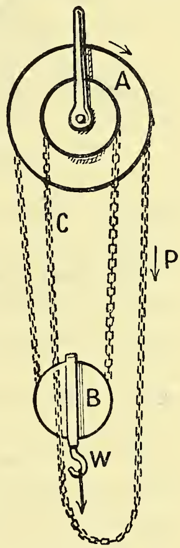


FIG. 34.

If R and r are the radii of the sheaves in block A, the amount of chain unwound when the sheave turns through an angle θ in the direction of the arrow is $R\theta$, while a length of chain $r\theta$ is wound up. Hence the length of chain which carries block B is reduced by an amount $(R-r)\theta$. Thus, while P moves through a distance $R\theta$, W moves through a distance $\frac{(R-r)\theta}{2}$. Hence neglecting friction and equating the work done by P to that done on W, we get

$$PR\theta = \frac{(R-r)}{2} W\theta$$

or

$$\frac{W}{P} = \frac{2R}{R-r} = v \quad . \quad . \quad . \quad . \quad (14)$$

when v is the velocity ratio. If when friction is taken into account

$P+x$ is the minimum force required to *raise* the weight W, then $P+x=W/v$, and the efficiency is

$$\frac{W}{(P+x)v}, \text{ or } \frac{W}{W-xv}$$

If the weight is being *lowered*, friction acts like a force of magnitude x in the direction of P, and it is evident that if x is greater than W/v, this frictional force will be sufficient to prevent the weight falling. If x is equal to W/v, the efficiency is $\frac{W}{W+W}$ or $\frac{1}{2}$. Thus if this form of tackle has an efficiency less than $\frac{1}{2}$, which as a matter of practice is always arranged for,[1] the weight will not be able to descend even when P is zero. To lower the weight in such a case a downward pull has to be exerted on the part C of the chain. This property of supporting the weight, even when no pull is exerted on the down pull, is of considerable convenience

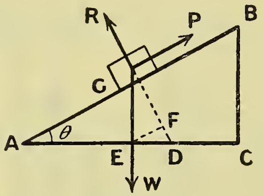


FIG. 35.

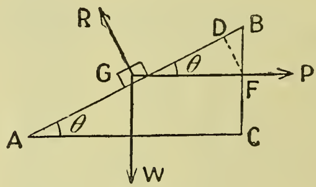


FIG. 36.

when the tackle is used to raise pieces of machinery, blocks of stone, and the like.

31. The Inclined Plane.—Suppose a body G (Figs. 35 and 36) rests on an inclined plane AB, and that there is no friction between the body and the surface of the plane, or at any rate that by suitable means friction is so much reduced as to be negligible in comparison with the other forces in play. The weight of the body W acts vertically downwards, and the reaction, R, of the plane acts perpendicular to the surface AB, so that if the body is to be prevented from sliding down the plane under the influence of the resultant of these two forces, it must be acted upon by a third force. The two principal cases which occur are when this force acts along a direction parallel either to the hypotenuse AB (Fig. 35) or the base AC (Fig. 36). In both cases the most convenient method of obtaining the relation between the force P, which we may call the effort, and W, for any given plane, is to use the principle of work.

Suppose G to move, under the influence of P, from A to B. Then the work done by P is P.AB, if P acts parallel to the hypotenuse, and P.AC if P acts parallel to the base, for in the one case the point of application

[1] Actual tackles have an efficiency of about ·3.

has moved through a distance AB in the line of action of the force, while in the other case the component of the displacement in the line of action is AC. In both cases the work done on W is $W \times$ CB, since the component of the movement of the point of application of W, parallel to its line of action, is CB. As in both cases the point of application of R moves at right angles to its line of action, no work is done on or by R during the displacement. Hence we have, when P acts parallel to AB,

$$P.AB = W \times CB$$

$$\frac{P}{W} = \frac{CB}{AB} = \sin\theta \quad . \quad . \quad . \quad (15)$$

When P acts parallel to AC,

$$P.AC = W \times CB$$

$$\therefore \quad \frac{P}{W} = \frac{CB}{AC} = \tan\theta \quad . \quad . \quad . \quad (16)$$

These results may also be obtained by resolving W parallel to the direction of P and equating this resultant to P.

Effect of friction with the inclined plane.

We have now to consider the effect of friction between the body and the surface of the inclined plane. First consider the case when the applied force P is parallel to the surface, *i.e.* AB (Fig. 34). Since the component of the weight W normal to the surface is $W\cos\theta$, the friction when motion takes place is $fW\cos\theta$. Hence the minimum force which will cause the body to move up the plane is $W\sin\theta + fW\cos\theta$, or $W(\sin\theta + f\cos\theta)$. On the other hand, if P is less than $W\sin\theta - fW\cos\theta$, the body will slide down the plane. For all values of P between these limits the body will remain at rest. For values of θ less than that for which $\sin\theta - f\cos\theta = 0$, or $\tan\theta = f$, that is, for inclinations of the plane less than the angle of friction, P must be negative to cause motion down the plane, or, in other words, P must act *down* the plane.

To form an idea as to the magnitude of the effect of friction, we may take the case of a smooth metal block sliding on a smooth metal plane, the surface being greasy, in which case f will be about 0·1. If the inclination of the plane is 15°, we shall have for the minimum value of P required to move the block up the plane, $W(\cdot259 + \cdot097)$, so that the increase in the value of P due to the friction is about 3·5 per cent.

When the body is moved through a distance x along the plane, the work done by the effort P is Px or $Wx(\sin\theta + f\cos\theta)$. During this movement the body is raised through a distance $x\sin\theta$, and thus the work done is $Wx\sin\theta$ (in gravitation units). Hence the *efficiency* of the arrangement, that is the ratio of the work done on W to the work expended by the effort, is $\dfrac{\sin\theta}{\sin\theta + f\cos\theta}$. Thus in the example above

the efficiency is $\frac{\cdot 259}{\cdot 259+\cdot 097}$ or ·73. If f were ·2, which is about the value when no grease is used, the efficiency would only be ·57.

As will be shown later, the difference between the work done by the effort and the useful work done on the body appears as heat at the surface where the friction occurs.

In the case where the force is parallel to the base, this force has a component normal to the inclined surface of $P \sin \theta$. Hence the frictional force just before the body starts is $f(W \cos \theta + P \sin \theta)$. The force parallel to the direction of P which will have this component parallel to the inclined surface is obtained by dividing this expression by $\cos \theta$. Hence, since the force when there is no friction is $W \tan \theta$, we have

$$P = W \tan \theta + f(W + P \tan \theta)$$

$$P = \frac{W(\sin \theta + f \cos \theta)}{\cos \theta - f \sin \theta} \quad . \quad . \quad . \quad . \quad (17)$$

It will be observed that the effect of friction is in this case much more important than when the force is parallel to the inclined surface. For all values of θ greater than that for which $\cos \theta - f \sin \theta = 0$, or $\cot \theta = f$, there can be no motion, however great the value of P.

Motion on an inclined plane.

Since the component of the weight of a body resting on an inclined plane acting down the plane is $W \sin \theta$, where θ is the inclination of the plane to the horizontal, if there is no friction, we shall have a force of $Wg \sin \theta$ absolute units acting on a body of mass W, and hence the acceleration, a, produced is given by

$$a = \frac{\text{Force}}{\text{Mass}} = g \sin \theta = g \times \frac{\text{Height}}{\text{Length}}.$$

Thus with an incline in which the height is half the length the acceleration will only be half that of a body falling freely.

If the coefficient of friction between the body and the plane is f, the frictional force will be $fWg \cos \theta$, and hence the force producing motion will be $Wg \sin \theta - fWg \cos \theta$, and the acceleration will be

$$g(\sin \theta - f \cos \theta).$$

If there is no acceleration, that is, if the inclination θ is so chosen that the body when once started continues to move down the plane at a constant speed, $\sin \theta - f \cos \theta$ must be zero, or $f = \tan \theta$; in other words, the inclination of the plane is the angle of repose (§ 24).

If a bicyclist coasts down a slope of 1 in x at a constant speed v, the weight of the rider and machine being W, the resistance to his motion at this speed must be equal to the component of W down the

slope, that is W/x, and the power expended against the resistance is Wv/x.

32. The Screw.—If a right-angled triangle ABC (Fig. 37) cut out of paper be wrapped round a cylinder, so that the base AB of the triangle lies entirely in a plane at right angles to the axis of the cylinder, then the hypotenuse AC will trace out a spiral line *aidfgc* on the surface

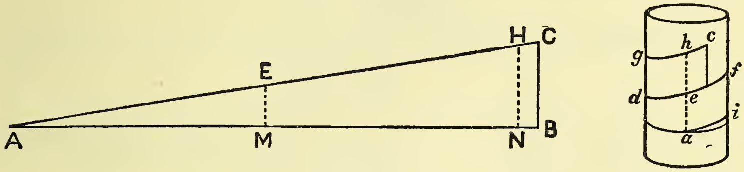


FIG. 37.

of the cylinder. If AM is equal to the circumference of the cylinder, then the distance, measured parallel to the axis of the cylinder, between the point a, which corresponds to the apex A of the triangle, and the point e, where a line drawn through a parallel to the axis meets the spiral line, is equal to EM. Hence if starting at a we follow the spiral line for a complete turn, we shall move along the cylinder parallel to the axis through a distance ae equal to ME. If we go twice round we shall move through a distance ah equal to NH, where MN is equal to AM. But since AN is double AM, NH must be double ME. Hence ah is double ae, so that for every complete turn the spiral line advances parallel to the axis through an equal distance. If a projecting ridge were fixed to the outside of the cylinder along the spiral line we should have a *screw*, the projection forming the *thread*. The distance between two consecutive threads is called the *pitch* of the screw. If a hollow cylinder has a groove cut on its inside surface so as just to fit the screw, it forms a *nut*. If the nut is turned through 360°, or one whole turn, it will move along the screw through a distance equal to the interval between two consecutive threads, or to the pitch of the screw.

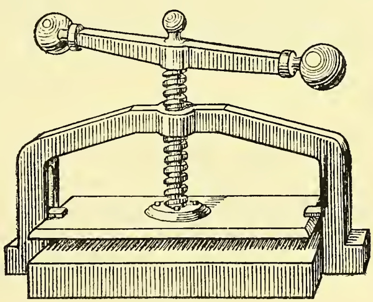


FIG. 38.

If, as in a screw press, we have a force P acting at right angles to the end of a cross arm (Fig. 38) of length $2R$ attached to a screw of which the pitch is h, so that the distance of the point of application of P from the axis of the screw is R, then during a complete turn the point of application of P will move through a distance equal to the

circumference of a circle of radius R, that is, $2\pi R$; and hence, since P is supposed always to act at right angles to the cross arm, the work done by P is $2\pi RP$. If W is the force exerted by the screw parallel to its axis, then the work done on W during a complete turn is hW, since the point of application of W will have been driven back through a distance equal to the pitch of the screw. Hence, if there were no friction,

$$2\pi RP = hW$$

or

$$\frac{P}{W} = \frac{h}{2\pi R} \quad . \quad . \quad . \quad . \quad (18)$$

For a given value of P we see that W is increased by decreasing the pitch of the screw, *i.e.* by having more threads to the inch, and by increasing the leverage R at which P acts.

In the case of a screw with a square thread as shown in Fig. 37, we can apply the results obtained for the inclined plane in order to estimate the effect of friction. If r is the radius of the screw, the equivalent inclined plane has a height equal to the pitch h, and a base equal to the circumference, $2\pi r$, of the screw, so that the inclination is given by $\tan\theta = h/2\pi r$.

If P is the force parallel to the base required to move a weight W up the plane, we found

$$P = \frac{\tan\theta + f}{1 - f\tan\theta} W \quad \text{(Equation 17, § 31)}$$

$$= \frac{h + 2\pi rf}{2\pi r - fh} W.$$

The moment of this force about the axis of the screw is Pr, so that the force Q acting at a distance R from the axis required to turn the screw is such that $QR = Pr$,

or

$$Q = \frac{Wr}{R}\left(\frac{h + 2\pi rf}{2\pi r - fh}\right) \quad . \quad . \quad . \quad (19)$$

To obtain an idea of the effect of friction, let us assume $f = \cdot 1$, $h = \cdot 1$ inch, and $r = \cdot 5$ inch. The force which acting at a distance r would turn the screw against an axial force W is

$$\frac{\cdot 1 + \cdot 1\pi}{\pi - \cdot 01} W, \text{ or } \cdot 132W.$$

If there were no friction the value would be

$$\frac{\cdot 1}{\pi}, \text{ or } \cdot 032W.$$

Thus friction increases the effort required about four times. To calculate

the efficiency, we have that the work done by the agent during one turn is $\cdot 132\,W \times 2\pi r$, and the work done on W is $\cdot 1\,W$. Hence the efficiency is

$$\frac{\cdot 1\,W}{\cdot 132\,W \times 2\pi \times \cdot 5}, \text{ or } \cdot 24.$$

The calculation in the case of an ordinary triangular thread is more complicated, and leads to a value of the efficiency about 3 or 4 per cent. less than in the case of a square thread of equal pitch and diameter.

CHAPTER VI

MEASUREMENT OF LENGTH, MASS, AND TIME

33. Measurement of Length. The Vernier. The Screw Micrometer. The Lever.—In the measurement of nearly all physical quantities, what we actually *observe* is the ratio of some length to some other length. Thus when we measure the pressure of the atmosphere with a mercury barometer, what we really observe and measure is the length of a column of mercury; the same statement applies to the measurement of a temperature with a mercurial thermometer; so also, when we use a spring-balance to measure a mass, it is the movement of a pointer along a scale that is observed. Hence we see the importance of being able to make accurate measurements of length.

With an ordinary scale divided into tenths of an inch it is possible, with a little care and practice, to measure by eye a length, which is not

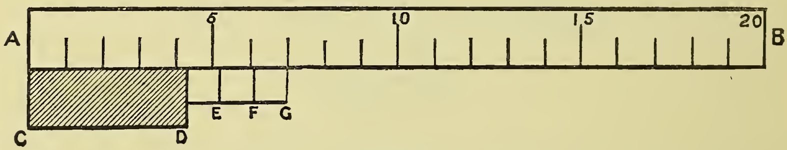


FIG. 39.

greater than that of the scale, to within one-hundredth of an inch. This is done by mentally supposing each of the tenths of an inch subdivided into ten equal parts, *i.e.* into hundredths of an inch, and estimating by eye by how many of these imaginary hundredth of an inch divisions the length exceeds the nearest number of whole divisions. In the same way, with a scale divided into millimetres, it is possible to read to tenths of a millimetre. In order to attain to a degree of accuracy much greater than the above it is necessary to adopt some mechanical means of subdividing the divisions, for merely making the divisions of the scale nearer together does not advance matters much if we trust to our judgment and eye alone, even if a magnifying glass is used. Of such mechanical contrivances the most commonly employed are the vernier and the micrometer screw.

The vernier consists of a small auxiliary scale, by means of which the divisions of the main scale can be subdivided without having to

use one's judgment as described above. The principle on which it works is as follows: Suppose AB (Fig. 39) is a scale divided into equal parts, and that the end D of some object (CD), the length of which is to be measured, lies between the fourth and fifth divisions, and that we require to find the fraction (x) of a division by which the length of CD exceeds four divisions of the scale. If a block DE, of such a length that ten of these blocks placed end to end would be equal to nine divisions of the scale (*i.e.* the length of each block is $\frac{9}{10}$ or 0·9 of a division), is placed at the end of CD, then the end (E) of this block will evidently project beyond the fifth division by an amount $(x-\frac{1}{10})$ of a division, since DE is $\frac{9}{10}$ of a division. If a second block EF is placed at the end of the first, the end F will exceed the sixth division of the scale by an amount $(x-\frac{2}{10})$ of a division. In the same way the end of a third block would project beyond the seventh division of the scale by an amount $(x-\frac{3}{10})$ of a division, and so on. It will be seen, however, that the end G of the third block exactly coincides with the seventh division of the scale, so that the amount by which it projects is zero. Hence

The vernier.

$$x-\tfrac{3}{10}=0,$$

or

$$x=\tfrac{3}{10}$$

That is, the length of CD is $4\frac{3}{10}$ or 4·3 divisions of the scale. We notice that if each of the blocks is $\frac{9}{10}$ths of a division in length, the object CD exceeds four divisions by as many tenths of a division as it is necessary to add blocks, till the end of the last block just coincides with one of the divisions of the scale.

It will be found that the above relation between the number of the blocks and the excess (x) is quite general, and it is utilised to mechanically subdivide the smallest divisions of a scale. Instead of having a number of separate blocks, it is more convenient to have a small auxiliary scale, called a vernier, which can slide along the edge of the chief scale, and is divided so that ten divisions of the vernier are equal to nine divisions of the scale. In this case we set the end of the vernier against the end of the object, and look along till we come to the division of the vernier which coincides with one of the divisions of the scale. In the case shown in Fig. 40, this occurs at the seventh division of the vernier, and hence the object is $4\frac{7}{10}$ or 4·7 divisions in length.

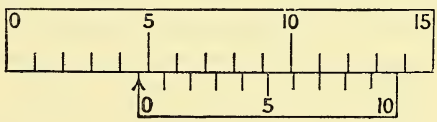


FIG. 40.

We may generalise and say that if m divisions of the vernier are equal in length to $m-1$ divisions of the scale, and coincidence occurs at

the nth division of the vernier, then the reading of the vernier, *i.e.* the distance between the zero line of the vernier and the preceding division line on the scale, is $\frac{n}{m}$ths of a division of the scale. Thus if the scale is divided into millimetres, 20 divisions of the vernier being equal to 19 mm., and coincidence occurred at the 17th division of the vernier, the reading would be $\frac{17}{20}$ mm.

Verniers are sometimes constructed so that m divisions of the vernier are equal to $m+1$ divisions of the scale. For instance, suppose ten divisions of the vernier are equal to eleven divisions of the scale. Then, by an argument exactly similar to that adopted above, it is evident that if the nth division of the vernier, counting as before from the end next the object which is being measured, coincides with a division of the scale, we have

$$x+\frac{n}{10}=1 \text{ division,}$$

or

$$x=\frac{10-n}{10} \text{ of a division.}$$

But $10-n$ is the number of divisions of the vernier between the coincidence and the end remote from the object. Hence if the vernier is numbered in the reverse direction to that in which the scale is numbered, the reading on the vernier will give directly in tenths the fraction of a division. The advantage of this form of vernier is that it is a little more open, *i.e.* the divisions are further apart, than in the other form.

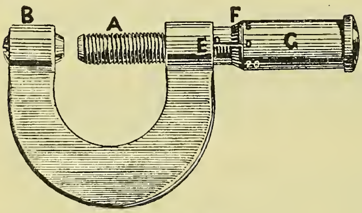


FIG. 41.

The micrometer screw.

In the micrometer screw use is made of the fact that if a screw is rotated through a complete turn its point will move, with reference to the nut, through a distance equal to the pitch of the screw, *i.e.* to the distance between two consecutive threads. By making the pitch of a screw small, and also attaching a drum-shaped head of considerable diameter which is divided into a number of equal parts, so that a fraction of a rotation can be read, it is possible to measure with great accuracy the distance moved over by the point of the screw.

An example of a case where a micrometer screw is used to measure a length is afforded by the screw-gauge shown in Fig. 41. The object to be measured is placed between the end of the screw A and the block B, which is connected by a strong curved arm with the nut in which the screw works. The number of whole turns made by the screw is read by

means of a scale, E, attached to the nut, which is gradually uncovered by the movement of the cap, G, attached to the screw. The fractions of a turn are read off on a scale, F, on the edge of this cap. The pitch of the screw ordinarily employed is 0·5 mm., and the edge of the cap is divided into 50 parts. Hence as turning the screw through a whole turn or 50 divisions advances the point A by 0·5 mm., one division on the scale F corresponds to a motion of the point of 1/50 of 0·5 mm. or 0·01 mm.

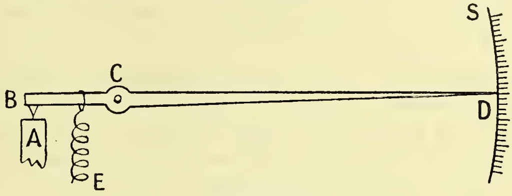


Fig. 42.

Magnifying lever.

A lever is occasionally employed to magnify the motion of some portion of an instrument, and hence allow of this motion being more easily measured. Suppose we require to measure the vertical movement of the top of the rod A (Fig. 42). If a lever BD, pivoted at C, has a point at the end of the short arm BC held against A by a spring E, then the movement of A will cause the lever to rock about the fulcrum C, and the end D of the long arm will move through a greater distance than does the end of A, and this magnified movement can be read by a scale S. If the movement as read off on S is x, the movement of A will be $x \times \frac{BC}{CD}$ by the principle of the lever (§ 27). To obtain a large magnification BC must be short and CD long. It is generally impracticable to make BC very small, as then the range of motion of A which can be dealt with is very small. On the other hand CD cannot be made very long, or trouble will be caused by the bending of this long arm. For this reason, when great magnification is required, the arm CD is replaced by a beam of light, the arrangement being called an optical lever.

Optical lever.

Thus suppose the long arm is replaced by a concave mirror M (Fig. 43), and a beam of light is thrown on the mirror along OM, this light being reflected by the mirror so as to give a spot on the scale S at F. When the lever rocks, owing to the movement of A, say downwards, the mirror will turn and the reflected ray will travel along MF′ so that the spot travels along the scale to F′. It will be shown later (§ 109) that the angle

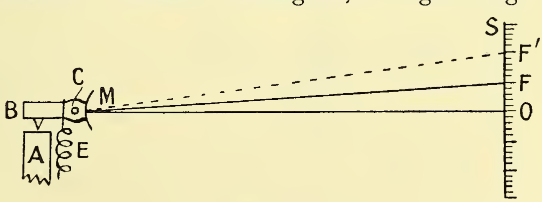


Fig. 43.

through which the reflected ray turns is twice the angle through which the mirror turns. Hence the movement of the spot will be twice as great as would be that of a material pointer of length CF, and hence if y is the movement of the spot on the scale, the movement of A will be $y \times \frac{FF'}{2CF'}$. As the beam of light has no weight and will not bend, the arm MF can be made very long, while as mentioned above the use of the mirror in itself doubles the movement, so that very great magnification can be obtained in this way.

34. The Measurement of Mass. The Balance. — Masses are compared by comparing the force with which they are attracted by the earth; in other words, by comparing their weights. The ordinary balance, which is the instrument most commonly employed for comparing weights, consists essentially of a lever of the first kind, called the beam, of which the arms are of equal length. The fulcrum consists of a knife-edge made of agate or steel attached to the beam resting on an agate or steel plane. The masses to be compared are supported in pans which are suspended from small planes resting on knife-edges attached to the beam. In a good balance beam the central and terminal knife-edges are all parallel and lie in the same plane. Let A and C (Fig. 44) be the points of support of the scale-pans, formed by the terminal knife-edges, and B be the fulcrum, formed by the central knife-edge, while BF is a pointer attached to the beam and by means of which the deflection of the beam can be read off on a scale S. Let w be the weight and G the centre of gravity of the beam and pointer. Then if the beam is to remain horizontal when the scale-pans are removed, the centre of gravity, G, must be vertically below the central knife-edge. If the lengths of the arms of the beam are equal, if the beam is to be horizontal when there are no weights in the pans, the weights of the pans must be equal. This being the case, if the knife-edges are all in one plane we may neglect the effect of the weight of the pans, since they will always have equal and opposite

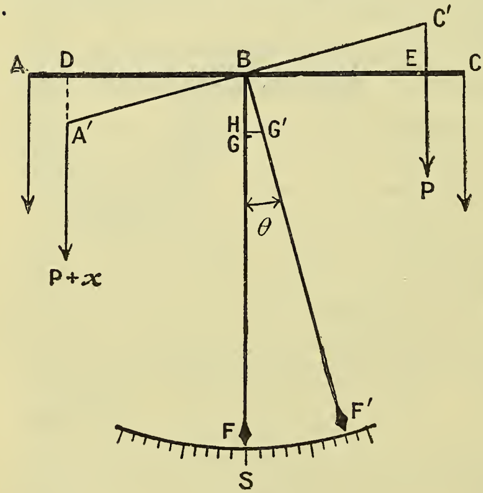


Fig. 44.

moments about the fulcrum. Suppose that in the right pan there is a weight P and in the left-hand pan a weight P + x, where x is a small quantity, so that the beam is deflected into a position such as A′BC′, making an angle θ with the horizontal. In order to find the connection between the excess x of weight in one pan and the deflection it produces we take moments about the central knife-edge. Thus on the left we have (P + x)DB. On the right the moment of the weight pan is P.BE, while since the centre of gravity of the beam is now at G′, the weight of the beam has a moment of amount w.HG′. If h is the distance BG of the centre of gravity of the beam below the central knife-edge when the beam is horizontal, BG′ = h and HG′ = $h \sin\theta$. Hence the total moment tending to turn the beam in the clockwise direction is P.BE + $wh \sin\theta$.

If a is the length of each arm of the beam, DB = $a \cos\theta$ = BE. Hence, making use of this relation and equating the moments, we get

$$(P+x)a\cos\theta = Pa\cos\theta + wh\sin\theta,$$

or

$$xa\cos\theta = wh\sin\theta.$$

$$\therefore \qquad \tan\theta = \frac{a}{wh}.x \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Now the greater the value of $\tan\theta$ the greater is the angle through which the beam has been turned by an excess of weight x in one pan, *i.e.* the greater sensitiveness of the balance. That is, the sensitiveness increases as the value of the fraction a/wh increases, and this increase can be produced either by increasing a or decreasing w or h. Hence, as far as sensitiveness is concerned, it is an advantage to use (1) a long beam (a large), (2) a light beam (w small), and (3) to arrange that the centre of gravity of the beam is only a very small distance below the central knife-edge (h small). The first two of these conditions are to a certain extent difficult to satisfy simultaneously, for if the beam is made very long, in order to be sufficiently stiff it has to be made heavy. A more serious matter, however, is that increasing the length of the beam increases the moment of inertia (§ 19) of beam and the load it carries about the central knife-edge, and this, together with any decrease in h, will cause the beam to swing very slowly when disturbed from its equilibrium position. The reason for this will be apparent when the subject of the time of vibration of a pendulum has been discussed. Since a very slowly swinging beam involves great loss of time in taking a reading, a compromise is generally advisable in which some sensitiveness is sacrificed by keeping the beam short, and h is not made very small. To compensate for this loss of sensitiveness to a certain extent a *long* pointer is attached to the beam, so that a small movement of the beam causes the end of the pointer to move over a considerable distance on the scale S (Fig. 44). Where great sensitiveness is required,

Sensitiveness.

the device used in the optical lever described in the last section is employed; a mirror being attached to the beam and a very long beam of light used as a pointer.

In the above discussion we have assumed (*a*) that the knife-edges all lie in one plane, and (*b*) that the arms are of exactly the same length, and we have now to consider what will be the effect of a slight departure from these ideal conditions.

Suppose ABC (Fig. 45) is a horizontal line through the central knife-edge, and that the end knife-edges are below this line at D and E. Then if each pan with its load exerts a vertical force W, and the distances between the end knife-edges and the central knife-edge are equal, the beam will be in equilibrium. If the beam is now displaced into the position $A'BC'$, it will be seen that the arm at which the force W at D' acts is less than the arm of the force W at C'. Hence the moments of these two equal forces about B are not equal, and the resultant moment tending to bring the beam back into its equilibrium position is equal to W multiplied by the difference between these arms. Therefore, since this moment depends on W or the load in the pans, the greater W is, the greater will be the tendency for the beam to return to its equilibrium position, so that it will require a greater *difference* in the loads in the two pans to deflect the beam.

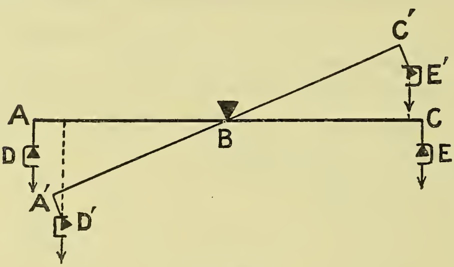


FIG. 45.

On the other hand, if the terminal knife-edges lie above the central knife-edge the sensitiveness will increase with the load, and it may even happen that the beam becomes unstable (§ 23), and hence the balance unworkable, for although in theory a condition of unstable equilibrium is realisable, this is not the case in practice.

Effect of inequality in length of arms.

If the arms are not of equal length, but the one on the left has a length a and that on the right a length b, the balance will be in equilibrium and the beam horizontal when

$$(P+S_1)a=(Q+S_2)b,$$

where S_1 and S_2 are the weights of the scale-pans. If S_1 and S_2 are so adjusted that the beam is horizontal when there are no weights in the pans,

$$S_1a=S_2b,$$

and hence

$$Pa=Qb \quad . \quad . \quad . \quad (a)$$

If Q is the weight of the body being weighed and P that of the weights, to obtain the value of Q we must multiply the value of the weights by a/b, that is by the ratio of the lengths of the balance arms. If the body is placed in the other pan and P' is the new value of the weights required to produce equilibrium, we have

$$Qa = P'b \quad . \quad . \quad . \quad . \quad (b)$$

Dividing (a) by (b)

$$\frac{P}{Q} = \frac{Q}{P'}$$

or

$$Q = \sqrt{PP'} \quad . \quad . \quad . \quad . \quad (21)$$

Thus by weighing the body first in one pan and then in the other and taking the square root of the products of the weights, the true weight independent of any equality of the lengths of the balance arms is obtained.

If, as is always the case, P and P' are very nearly equal, and we call their difference x, we have

$$\sqrt{PP'} = \sqrt{P(P+x)} = P\sqrt{1+\frac{x}{P}}$$

But since x/P is very small, we can expend $\sqrt{(1+x/P)}$ by the binomial theorem and neglect terms involving the square and higher powers of the small quantity x/P. Hence

$$\sqrt{PP'} = P\left(1+\frac{x}{2P}\right) = \frac{2P+x}{2} = \frac{P+P'}{2}$$

Thus when the difference between the length of the arms is small we may obtain the true weight of a body by taking the mean of the weights required to produce equilibrium when the body is placed first in one pan and then in the other.

35. Time. Solar and Mean Time. — The measure of time is based on the rotation of the earth; and the second, which is the scientific unit of time, is the 86,400th part of a *mean solar day,* that is, of the *average* interval which elapses between successive transits of the sun across the meridian at any place during a whole year.

Owing to the eccentricity of the earth's orbit, and the fact that the earth's axis is not perpendicular to the plane of the orbit, the interval between two successive transits varies during the year, so that the actual solar day is not the same as the mean solar day.

If a clock keeps mean time and agrees with solar time, that is time such as would be indicated by a sundial, when the sun appears in that portion of the heavens known as the first point of Aries, then the difference between the time of noon as indicated by this clock, and

the time when the sun crosses the meridian on any day, is called the equation of time at noon for that day. The curve in Fig. 46 gives the equation of time for the year. When the curve is above the axis OX the equation of time is positive, that is, the time as shown by a mean-time clock will be ahead of the transit of the sun by the amount shown by the ordinate.

Equation of time.

It will be seen that the equation of time is zero, that is, the time as indicated by a mean-time clock and the sun will be the same on April 15, June 15, August 31, and December 24. On February 11 the mean-time clock is 14 minutes 29 seconds ahead of the sun, while on November 1 it is 16 minutes 20 seconds behind the sun.

The unit of time used in astronomy is the sidereal day. This represents the interval between two consecutive transits of one of the fixed stars across the meridian. Since the distance between the earth

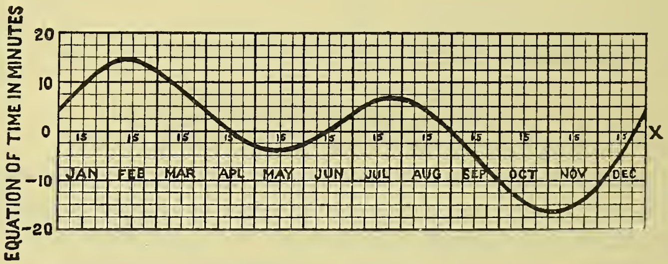


Fig. 46.

and any of the fixed stars is very great, compared even with the diameter of the earth's orbit, the line joining the earth to such a star remains always parallel to itself. Hence the sidereal day represents the time the earth takes to make one complete rotation about its axis. A sideral day is equal to 23 hours 56 minutes 4·09 seconds of mean solar time.

The use of the rotation of the earth as a measurer of time is not without objection, for there can be no doubt that the mean solar day is gradually growing longer, due to the slowing down of the rotation of the earth. In order to remove this objection, it has been proposed to use the time of vibration of the atom of some element, such as sodium, as the unit of time, for under definite conditions it appears as if this time were quite fixed and unalterable.

Clocks and watches depend for their action on the fact that a pendulum and a balance-wheel, subject to certain restrictions, take a definite time to execute a vibration, and consist of a mechanism by which enough energy is supplied to the pendulum or balance-wheel to allow for losses due to friction, so that the motion may be main-

tained, and an arrangement for counting and indicating the number of vibrations. The energy is generally supplied by a raised weight or a bent spring. The conditions that a pendulum may always complete a vibration in the same time will be considered later (§ 86).

The sundial.

An instrument in which the sun itself is used to indicate the time is the sundial. Even a comparatively roughly made sundial will allow of the correct time being obtained to within about a minute, and such a dial has considerable practical value in the country, where it is not always easy to obtain the "correct time" by which to set ordinary watches and clocks. The principle on which the sundial works is as follows: Suppose we imagine the earth fixed and the sun to rotate, then if a line ON (Fig. 47), called the polar axis, is drawn parallel to the axis of the earth, the path of the sun will be a circle perpendicular to ON. Hence if we had a circle BCD fixed perpendicular to ON, at noon the shadow of ON would cut this circle at C, and the point where the shadow cuts the circle would travel round the circle at a uniform rate equal to a whole turn in 24 hours, or of 15° per hour. Thus if the circle is divided into 15° intervals, representing hours, and 15 minutes of arc, representing minutes of time, starting from C as noon the *solar* time can be read off from the position of the shadow. Such a circle would, however, be impracticable since it would itself cause a shadow. This difficulty is got over by having a *horizontal* circle ECF with its centre at O, which in place of being uniformly divided as was the circle BCD, is divided so that corresponding to the point D we have a point F, such that FOAD are all in one plane. Then when the shadow of the rod would cut the circle BCD at D, it will cut the horizontal circle ECF at F, and hence this circle can be graduated to read in hours and minutes. The time as given by a sundial is of course solar time, and to obtain mean time the equation of time has to be applied. Even then the time obtained is local time, and to obtain standard time, say Greenwich time, an allowance of 15 minutes for each degree of longitude *west* of the standard place (Greenwich) has to be added to the local time. If the place of observation is east of the standard station, the correction for longitude must be subtracted.

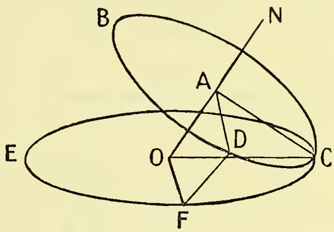


Fig. 47.

CHAPTER VII

GENERAL PROPERTIES OF MATTER

36. Elasticity.—When a system of forces act on a portion of matter, although they may not produce any motion of the body as a whole, yet they may produce a displacement of the various particles of the body relatively one to the other. Such a relative motion of the different portions of a body is called a *strain*. In the preceding pages, when dealing with the action of forces on a body, we have supposed the body to be rigid; that is, we have supposed that, however great the forces acting on the body, they did not produce any strain.

When a body is strained by the application of external forces it in general resists the strain, forces being called into play by the relative displacement of its parts tending to cause it to return to its unstrained condition. This restoring force, called into play owing to the strain, is called a *stress*.

Strain and stress.

It is important to note that the magnitude of the stress induced in a body is not necessarily always equal to that of the forces applied to produce the strain. Thus it requires a considerable force to deform or strain such a body as a lump of moist clay; but if the external forces are removed, there is no tendency for the clay to regain its former shape. If, however, the deforming forces strain the body till the stress induced is such as just to prevent further strain, then a state of equilibrium will be set up, and the stress will be exactly equal and opposite to the deforming force.[1]

A body which offers a stress, tending to restore the body to its original condition when it is strained, is said to be elastic or to possess elasticity.

The elasticity of a body is *measured* by the ratio of the stress corresponding to a given strain to that strain, or

$$\text{Elasticity} = \frac{\text{stress}}{\text{strain}} \quad . \quad . \quad . \quad . \quad (22)$$

There are two distinct types of strain, namely, that in which the volume is altered without change of shape, and the other in which the

[1] As we shall only deal with the cases where a strained body is in a state of equilibrium, the deforming forces will always be equal and opposite to the stress called into play. Hence it saves circumlocution if we use the term stress to indicate the deforming forces.

shape is altered without change in volume, and to which a special name, a *shear*, is given. Strains, however, often consist of a combination of the two types, so that both the volume and shape are altered.

In the case of a volume change the strain is measured by the change in volume produced per unit volume. Thus, if by the application of outside forces the volume of a body changes from V to $V-v$, so that the decrease in volume is v, the strain is v/V.

The simplest case of a shear is that in which a rectangular block of the material ABCD (Fig. 48) is deformed by the application of equal and opposite forces in the directions P and Q applied to the opposite sides AB and CD of the block, so that the block is distorted to the position A′B′CD. It is evident that if the forces P and Q were to act alone, the block being under the action of no other external forces such as its weight, then the block would simply rotate in place of being distorted. In order to prevent the rotation as a whole, another pair of equal parallel opposite forces P′, Q′ must of necessity be applied. Leaving, however, the consideration of this point aside for the moment, the strain produced is measured by the circular measure of the angle ADA′ or BCB′. When, as is almost invariably the case, this angle is small, the circular measure of ADA′ is AA′/AD. Hence the strain is equal to AA′/AD, and is equal to the relative displacement of the parallel plane faces AB and DC divided by their distance apart, or what amounts to the same thing, by the relative displacement of two planes at unit distance apart.

Simple shear.

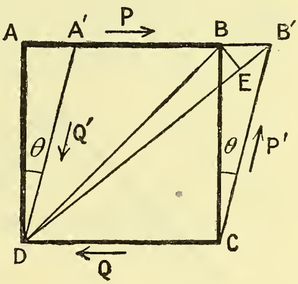


Fig. 48.

The kind of distortion which takes place when a body undergoes a pure shear may be illustrated by considering a pack of cards, first piled so as to form a rectangular block as at (*a*) (Fig. 49), and then displaced so that each card has been moved a small distance to the right with reference to the one below, as shown at (*b*). In the case of the cards, when displaced there is no tendency to return to the original formation. If we imagine that each card is connected to its neighbours by elastic, then when displaced there would be a force tending to restore the cards to their original place, and we should more nearly represent what occurs in an ordinary solid block.

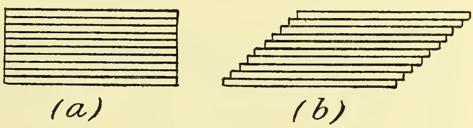


Fig. 49.

Returning to the case illustrated in Fig. 48, it will be observed that owing to the shear the diagonal DB′ has been increased and the diagonal A′C decreased in length. If we suppose that the body which is strained is a cube, each edge having a length a, then the shear θ is BB′$/a$. If BE is drawn perpendicular to DB′, the increase in length of the diagonal DB is sensibly equal to EB′. Denoting by s the increase in length *per unit length* we have $s = \text{EB}'/\text{DB} = \dfrac{\text{EB}'}{a\sqrt{2}}$. Since in all actual cases the angle θ is very small, the angle BB′E is sensibly equal to 45°, so that $\text{EB}' = \text{BB}'/\sqrt{2}$. Hence

$$s = \frac{\text{BB}'}{2a} = \frac{\theta}{2} \qquad . \quad . \quad . \quad . \quad (23)$$

In the same way it can be shown that decrease in length of the diagonal A′C per unit length is equal to $\theta/2$.

Next let us suppose that a cube MNOP (Fig. 50) is subjected to two inward directed systems of forces, or a compression, acting over the faces MN and OP, and to two outward directed forces, or a tension, acting over the faces NO and MP. As a result the cube will be deformed into a rectangular block of section M′N′O′P′. Now consider a square ABCD in the face of the block before it is strained. Owing to the strain the square will be distorted into the figure A′B′C′D′, that is, the diagonal AC will be shortened and the diagonal DB will be lengthened, just as occurred under the influence of the forces shown in Fig. 48. That is, the portion of the cube ABCD will have undergone a shear. This result can at once be extended to the whole of the cube, so that we infer that a uniform compression and a uniform tension at right angles will produce a shear, and it can be shown that if the original edges are each of unit length, then when the shear takes place the sides of the rectangle M′N′O′P′ are $1-s$ and $1+s$ respectively.

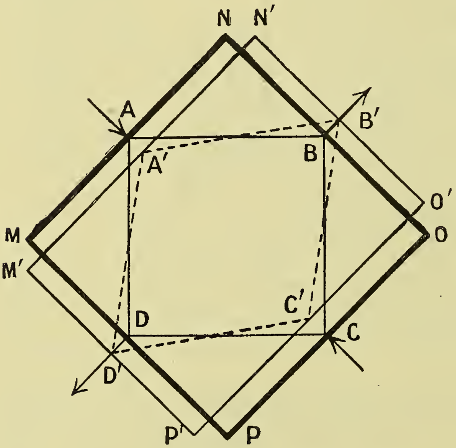


FIG. 50.

We have now to consider the stress set up when a body is strained. Since when the strain is complete the body is in a state of equilibrium,

it follows that the stresses set up owing to the strain must have a resultant equal and opposite to the resultant of the external forces which maintain the strain. Hence if we obtain the magnitude of these external forces, we can at once deduce the stress. For simplicity we shall only consider cases where the stress is uniform throughout the body.

Measure of a stress.

In the case of a pure volume strain, the force perpendicular to the surface acting on each unit of surface must be everywhere the same in magnitude, and this force per unit area measures the stress. We shall see later that a system of forces such as we have postulated above constitute what is called a hydrostatic pressure, and can be most easily applied by placing the body in a vessel containing a fluid, and then by means of a pump forcing more fluid into the vessel.

In the case of a shear we have seen that it can be regarded in two ways. In the first of these, illustrated in Fig. 48, there is a tangential force acting. Now it is obvious that the displacement produced by this force for any given material will depend on the height AD (Fig. 48) of the rectangular block, and also on the cross-section of the block taken perpendicular to the plane of the paper, that is, to the area of the face on which the tangential force acts. Hence for a given height of the block the displacement BB′ will be proportional to the tangential force per unit area.

It has been shown by experiment that so long as the strain is kept below a certain limit for each material, called the elastic limit, then the stress is proportional to the strain, and hence the ratio of stress to strain, that is the elasticity, is a constant. This result is known as Hooke's law, and we shall see later how it can be tested by experiment.

Hooke's law.

Let us now consider a cubical block of any given material of unit volume, and when the shear is θ let the tangential force be T. If we start with the block unstrained, so that the tangential force is zero, and gradually increase the force to T, the strain will increase in exactly the same proportion, and during the process the force will do work. Now it can be shown by an argument exactly similar to that given in § 17, that the total work done when the final shear is θ and the final stress is T is half the product of the final value of the force into the displacement of its point of application. The displacement is θ since the height of the block is unity, hence the work done is $T\theta/2$.

Next consider the second manner in which a shear can be produced, namely, that illustrated in Fig. 50. If s is the total shortening in one direction, and the original shape is a unit cube, the work done by the compressing forces is equal to half the product of the final value of these forces into the displacement, or if P is the final value of the force acting

on unit area, the work is $Ps/2$. In the same way the work done by the tensions P acting on the faces which are pulled out is $Ps/2$, and hence the total work is Ps. We have, however, seen that $\theta = 2s$, hence it follows, since the same amount of work must be done to produce a given strain in a given body, whatever the manner in which it is produced, that $P = T$. In other words, the pull and push P per unit area required to produce a given shear is equal to the tangential force T required to produce the same shear.

The elasticity, n, which is the ratio of the stress to the strain, and in the case of a shear is called the coefficient of *rigidity*, is given by

$$n = \frac{T}{\theta} \text{ or } \frac{P}{2s} \quad . \quad . \quad . \quad . \quad (24)$$

and since $T = P$ and $\theta = 2s$, the value is the same whichever method of producing the shear is considered.

37. Density.—Since all matter has mass, and also occupies space, it is important to consider the relation of the mass of any particular portion of matter to the space occupied. This relation is called the density of the body. Thus if a body of mass M is uniform throughout and occupies a space V, the density D is given by $D = M/V$, or in other words, the density is the mass of unit volume.

In the C.G.S. system the density of a substance is the mass in grams of a cubic centimetre. In the British system the density is the mass in pounds of a cubic foot.

Since the mass of a cubic centimetre of water at 4° C. is one gram (§ 2), it follows that the density of water at that temperature is unity in the C.G.S. system.

Since a cubic foot of water at 4° C. (39° F.) weighs 62·42 lbs., the density of water in the British system is 62·42.

The term specific gravity, or relative density, is often used to indicate the ratio of the density of a body to that of water. If the temperature of the water is always taken as 4° C., then in the C.G.S. system the specific gravity is numerically the same as the density. The specific gravity is the same for any given substance in both the C.G.S. system and the British system.

In the C.G.S. system the weight in grams of any body can at once be obtained by multiplying the specific gravity or density by the volume in cubic centimetres. The volume in litres multiplied by the specific gravity will give the weight in kilograms. In the British system the weight in pounds is equal to 62·42 times the product of the specific gravity into the volume in cubic feet, or to ten [1] times the product of the specific gravity into the volume in gallons.

[1] A gallon of water at 39° F. weighs 10 lbs., hence there are 6·242 gallons in a cubic foot.

38. States of Matter.—For the purposes of subdivision we may say that matter exists in three distinct states, the solid, the liquid, and the gaseous. In addition, however, to states which fulfil the definitions of a solid, a liquid, or a gas, which we shall give later on, it will be found that there are intermediate states which bridge over the intervals between the solid and the liquid, and the liquid and the gas. As an example of the kind of gradation which exists, we may take the following: steel, lead, wax, cobbler's-wax (which will flow like a liquid if allowed sufficient time), treacle, water, ether, liquefied carbon dioxide, steam, sulphur dioxide, air, hydrogen. In addition there is the critical state when a substance is to all intents and purposes both a liquid and a gas.

We may define a solid as a portion of matter which is able to support a steady longitudinal stress without lateral support. In contradistinction, a portion of matter which is unable to support a steady longitudinal stress without lateral support is called a *fluid*.

If we take a solid body, say a lead-pencil, then we may apply a deforming force, either of compression or extension, in any direction to the pencil, and there will be a certain amount of strain, either elongation, compression, or bending, produced, which will call into play a stress that, so long as the deforming force is not too great, will be in equilibrium with this force, and this stress will be produced without the body being supported in any way in a direction at right angles to that along which the stress acts. In the case of a fluid, such as water or air, we are unable to exert a stress on it, and hence produce a corresponding strain, unless we supply some constraining boundary which shall prevent the fluid swelling out at right angles to the line of action of the stress. Thus if we have a fluid enclosed in a cylindrical tube between two pistons, then we may apply a deforming force to the fluid either by forcing the pistons towards one another, or by pulling them apart, in one case producing a compression, and in the other a tension in the direction of the axis of the tube, and a stress will in both cases be produced in an opposite direction to the applied force. If, however, part of the wall of the tube between the pistons is removed, and we then attempt to apply stress to the liquid, we shall not succeed, for either the liquid will flow out through the gap in the tube, or air will be sucked into the tube through this opening, and the fluid will remain unstrained. It is only, therefore, when the column of fluid is laterally supported by the walls of the tube that it is capable of exerting a longitudinal stress.

Fluids are divided into liquids and gases. A liquid is a fluid such that when a certain volume is introduced into a vessel of greater volume it only occupies a portion of the vessel equal to its own volume. A gas is a fluid such that if a certain volume is introduced into a vessel, then,

whatever the volume of the vessel may be, the gas will distribute itself uniformly throughout the vessel.

39. Pressure Exerted by a Fluid. Principle of Archimedes.—Since a fluid cannot resist a stress unless it is supported on all sides, or in other words it has only elasticity of volume, it can offer no permanent resistance to forces which tend to change only its shape and not its bulk.

It follows, from this mobility of fluids, that in the case of a fluid at rest the force it exerts on any surface in contact with it must be perpendicular to the surface. If the force did not act perpendicular to the surface, then the *reaction* could be resolved into two components, one acting perpendicular to the surface, and the other acting parallel to the surface. This latter component would, if it existed, cause the fluid to move parallel to the surface. Since by supposition the fluid is at rest, and therefore no such tangential motion exists, there can be no tangential component of the force, so that the force exerted by the fluid on the surface is perpendicular to the surface.

The magnitude of the force exerted by a fluid is measured by the force exerted on the unit of surface, and is called the *pressure*.

If the pressure over a surface is not uniform, then we measure the pressure at a point by considering the force exerted on an element of area, taken round the given point, so small that the pressure is practically constant over this area, and divide the force by the area; a process exactly analogous to that adopted in the case of a variable speed.

It also follows, as a consequence of the mobility of fluids, that if we apply a pressure to a fluid enclosed in a vessel, then the fluid will transmit this pressure equally in all directions.

Fluids transmit pressure equally in all directions.

If a fluid is unacted upon by any other force besides the pressure of the sides of the containing vessel, then the pressure must be the same at every point within the fluid, and must act at every point equally in all directions. This statement may be proved by imagining that a small cubical element of volume of the fluid becomes solidified without any other change. This element will evidently still remain in equilibrium, and hence the forces acting on all the faces must be equal. As the area of all the faces is the same, this means that the pressure on all the faces must be the same. Since this must hold good however the small cube is turned, it follows that the pressure must be the same in all directions.

In a fluid at rest, and acted upon by gravity, the pressure in the lower layers is greater than in the upper, since each layer has to support the weight of all the superincumbent layers. The pressure throughout any horizontal layer must, however, be everywhere the same. Otherwise, if there were two points in the same horizontal plane at which the pres-

sure was different, then, since no work would be done against gravity by the passage of fluid from one of these points to the other, if we had a small pipe with one end at one point, and the other end at the other, the fluid would flow from the point of higher pressure to the point of lower pressure through the pipe. This motion would also take place even if no pipe connected the two points, and hence the fluid would not be at rest, which is contrary to hypothesis. If the two points are at different levels, then the pressure at the lower point is greater, but the liquid there does not move to the higher point, since, to do so, work would have to be done against gravity.

Principle of Archimedes.

When a solid body is immersed either partly or wholly in a fluid, it displaces a volume of the fluid equal to the volume of the immersed part, and it experiences an upward force, due to the fluid, equal in magnitude to the *weight* of the volume of fluid displaced. This is known as the *Principle of Archimedes*, and its truth may be proved as follows. Suppose that the body immersed in the fluid is a cube *abcdefgh* (Fig. 51), and that it is immersed so that the edges *ae*, *bf*, *cg*, and *dh* are vertical. Then the total pressure of the fluid on the face *adhe* is exactly equal and opposite to the total pressure on the face *bcgf*. For we may suppose them each divided into equal horizontal strips, so that the pressure is constant over each strip. Then for each strip in one face there is an equal strip in the other in the same horizontal plane, so that the pressure is the same. Hence the total pressure exerted on each pair of strips is equal and opposite; and therefore the total pressure on one face is equal to the total pressure on the other. The same argument applies to the faces *abfe* and *dcgh*. The upward pressure on the face *efgh*, since it is at a lower level, is, however, greater than the downward pressure on the face *abcd*. In order to see what is the difference between these two forces, suppose the cube removed and replaced by a cube of the given fluid, which by some means has been solidified without any other change. This cube will then be in equilibrium in the fluid. The total pressure on the vertical faces will as before exactly balance each other, so that this cube of the fluid is in equilibrium under the three following forces: (1) the weight of the cube of fluid acting downwards; (2) the total pressure of the fluid on the upper face *abcd* acting downwards, and (3) the total pressure of the fluid on the lower surface *efgh* acting upwards. These forces are all parallel, and so, in order that there may be equilibrium, the sum of the two downward acting forces must be equal to the upward acting force; that is, the difference between the total upward fluid pressure and the total downward

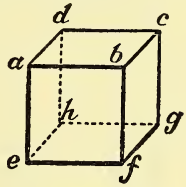


Fig. 51.

fluid pressure is equal to the weight of the cube of fluid. Hence when the solid is immersed in the fluid, since its upper and lower faces occupy just the same positions, as we have supposed the faces of the fluid cube to occupy, the difference in the total pressures[1] on the lower and upper faces will be the same as before, *i.e.* it will be equal to the weight of a cube of the fluid equal in size to the solid, or, in other words, the upward force acting on the immersed solid will be equal to the weight of the fluid displaced by the solid. This upward force acts through the centre of gravity of the *displaced fluid.*

[1] Where no ambiguity can be caused, it is usual to employ the word *pressure* where *total* pressure is really meant.

CHAPTER VIII

PROPERTIES OF GASES

40. Elasticity of Gases. Boyle's Law.—The only kind of elasticity of which a gas is capable is elasticity of volume or bulk elasticity, since it is only to a change of volume that a gas offers any permanent resistance. If the pressure acting on a volume V of a gas is increased from P to $P+p$, and as a result the volume becomes reduced to $V-v$; then the strain or deformation produced in a volume V is v, and therefore the strain produced per unit volume is v/V, while the stress corresponding to this strain is p. Hence, since the elasticity of a body is the ratio of the stress to the strain it produces, the elasticity e of the given gas is

$$e=p\div\frac{v}{V}=\frac{pV}{v} \quad . \quad . \quad . \quad (25)$$

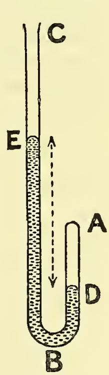


FIG. 52.

In order to study the elasticity of gases, Robert Boyle made use of a glass U-shaped tube (Fig. 52). The end of the shorter limb AB of this tube was closed at A, while the end of the longer limb was open. Having calibrated the shorter limb, so that the volume occupied by the gas enclosed in it was known, mercury was poured into the open limb, so as to imprison a certain quantity of air in the closed limb. The volume of the enclosed gas was obtained from the previous calibration of the tube, and the pressure to which it was subjected was the atmospheric pressure (§ 41), together with the weight of a column of mercury of height ED. By adding more and more mercury, and reading the corresponding values of the volume and the height ED, Boyle obtained a series of values of the volume of a given mass of air under different pressures, and as a result he was led to enunciate a law, which is known as Boyle's[1] law, that the product of the volume of a given mass of a gas into the pressure is a constant for all pressures.

Boyle's law.

If V is the volume of a given mass of gas, and P is the pressure to

[1] This same law was announced some fourteen years later by Mariotte, and is sometimes referred to as Mariotte's law in consequence.

which it is subjected, then Boyle's law states that so long as the temperature remains constant,

$$PV = \text{constant} \quad . \quad . \quad . \quad . \quad (26)$$

If the pressure acting on a gas is increased from P to $P+p$, and as a consequence the volume changes from V to $V-v$, then, if the temperature is kept constant, according to Boyle's law we have

$$PV = (P+p)(V-v)$$
$$= PV + pV - vP - pv$$

or

$$\frac{pV}{v} = P + p$$

But pV/v is the elasticity of the gas. Hence we see that the elasticity is numerically equal to the *final* pressure. If p is so small that we may neglect it compared to P, that is, if we measure the elasticity for a small range of pressure, we may say that the elasticity of a gas at a pressure P, and at *constant temperature*, is numerically equal to P, and is thus the same for all gases at this pressure. Since, as we shall see, the pressure of the atmosphere is about 10^6 dynes per square centimetre, the elasticity of a gas at this pressure is 10^6 dynes per square centimetre.

Elasticity of a gas at constant temperature is equal to the pressure.

Boyle only tested the accuracy of his law over a small range of pressures, and further his apparatus did not allow of the measurements being made with any degree of accuracy. Since his time his experiments have been repeated, notably by Regnault, Amagat, and Rayleigh. Regnault's experiments extended up to pressures of 27 atmospheres, Amagat's up to 3000 atmospheres, while Rayleigh confined himself to a very accurate test of the law at pressures up to half an atmosphere.

The results of these investigations has been to show that Boyle's law only holds exactly over a fairly small range of pressure. For large changes of pressure the product PV is no longer constant, the departure varying with the nature of the gas, the temperature, and the value of P. Some of Amagat's results are given in Fig. 53, where the values of the product PV for different values of P are plotted against the pressure. The value of PV at atmospheric pressure is in each case taken as unity.

Departures from Boyle's law.

If a gas obeyed Boyle's law exactly at all pressures it would give a horizontal line, and hence the departure of the curve for each gas from such a horizontal line through the point A indicates the amount of the departure from Boyle's law. In the case of hydrogen the value of PV increases steadily with pressure. With nitrogen the value of PV at first decreases slightly, and then

increases more rapidly than in the case of hydrogen. With oxygen the initial fall is more marked, but the curve in general resembles that for nitrogen. The above curves all correspond to a temperature of 0° C. The curve for carbon-dioxide for this temperature has an entirely different form, dropping to a sharp peak, which is due to the fact that the gas liquefies, the temperature being below the critical temperature (§ 80). At 100° C. the curve for carbon-dioxide resembles more nearly the curve for oxygen, while at 258° C. the dip is even less than that from oxygen

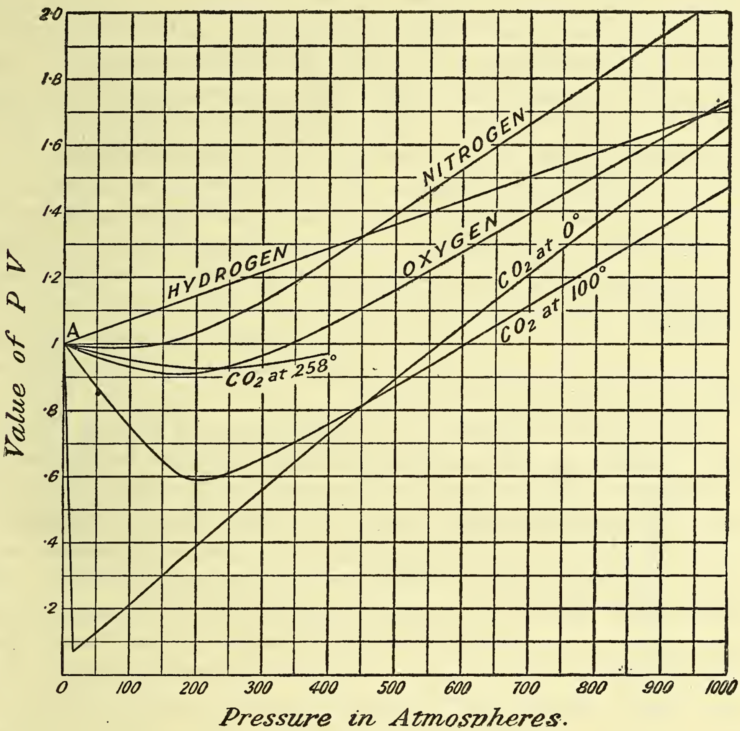


FIG. 53.

at 0°. It will thus be seen how, as the temperature increases, the easily liquefiable gas carbon-dioxide resembles more nearly one of the so-called permanent gases.

In the case of a gas, A, for which the product PV increases as P increases, the elasticity is greater than for a gas, B, for which PV is constant, so that Boyle's law holds exactly. This is at once evident, for to produce a given strain (decrease in volume) a larger increase in pressure or stress is necessary in the case of A, since the product of the

pressure into the volume is greater for A than for B, and by hypothesis the volumes are the same.

41. Pressure of the Atmosphere. The Barometer.—In the year 1643 an Italian named Torricelli discovered that if you take a glass tube closed at one end and about a metre long, fill it with mercury and then invert it so that the open end is below the surface of the mercury in a trough, the mercury no longer completely fills the tube. The height of the column of mercury is only about 76 centimetres or 30 inches. If the tube is inclined the mercury extends further along the tube so that the *vertical* height of the meniscus in the tube above the surface of the mercury in the trough remains always the same. As the inclination is increased the mercury finally completely fills the tube, and unless the inclination of the tube is performed very slowly and steadily the mercury hits the top of the tube with a characteristic click. Torricelli gave the true explanation of the above phenomenon, namely, that the mercury column is supported by the pressure of the atmosphere acting on the surface of the mercury in the trough, and that the pressure exerted by the atmosphere is equal to the weight of a column of mercury about 76 centimetres high. An instrument for measuring the atmospheric pressure is called a *barometer*, and the height of the mercury column in Torricelli's experiment is called the *barometric height.*

Barometric height.

The barometric height is not a constant even at a given place, but varies, sometimes quite rapidly, such rapid variation being generally accompanied by high winds. At places near sea-level and in temperate zones the barometric height is generally between 71 and 79 centimetres (28 and 31 inches). It will thus be seen that the term atmospheric pressure does not indicate a pressure which is even approximately constant. Thus the statement that the volume of a certain quantity of a gas measured at *atmospheric pressure* has a given value is very vague, for the volume of gases is largely affected by the pressure. Thus the volume of a given quantity of gas measured at a pressure of 75 centimetres of mercury would increase by 4 per cent. if the pressure fell to 72 centimetres. Except, therefore, in very rough measurements it becomes necessary to adopt some standard value for "atmospheric pressure." The standard pressure generally adopted is such that it will support a column of mercury 76 centimetres high, the temperature of the mercury being that of melting ice, and the attraction of gravity having the value which exists at sea-level and at latitude 45°. The reason why the temperature and the value of gravity have to be specified is that the pressure required to support a given column of mercury depends on the attraction of the earth on the mercury, and this depends on the density and the value of g.

The standard atmosphere.

Since the density of mercury at 0° C. is 13·596, and the value of g at latitude 45° and sea-level is 980·60 c.m./sec.^{2}, a standard atmosphere as defined above is equal to $76 \times 13·596 \times 980·60$, or 1013250 dynes per square centimetre. Since this number is very nearly a million or 10^6, it has been proposed to take a million dynes per square centimetre as the standard atmosphere, and this pressure is sometimes called a metric atmosphere. It corresponds to a column of mercury of 75·005 centimetres height, the temperature being 0° C., and the measurement being made at sea-level and latitude 45°.

Barometers may be divided into two classes: (1) Liquid barometers, in which the pressure is measured in terms of the height of a column of a liquid, and (2) aneroid barometers, in which the pressure is measured by the deformation of the lid of a metal box.

The barometer.

Practically the only liquid that is used in liquid barometers is mercury, since, on account of its great density, the height of the column which the pressure of the atmosphere can support is of a manageable magnitude. Another advantage possessed by mercury is that it does not absorb moisture from the air, as does glycerine, the only other liquid that has been used to any extent.

The simplest form of mercury barometer is the syphon barometer. It consists of a U-shaped tube, the longer limb (AB, Fig. 54) of which is closed at A, and is about 86 centimetres long, while the shorter limb is open at C. This tube is filled with pure mercury, and by boiling the mercury any air or moisture adhering to the mercury or to the bore of the tube is expelled. The distance DE is equal to the barometric height. When the barometric pressure increases, the mercury rises in the closed limb and falls in the open limb; and if the bore of the two limbs is the same, the movement of the mercury surface (meniscus) is the same in the two limbs but in opposite directions. Hence, if the mercury rises in the closed limb by 1 centimetre, it will also fall in the open limb by 1 centimetre, and therefore the distance DE will increase by 2 centimetres, that is, the atmospheric pressure will have increased by 2 centimetres of mercury.

FIG. 54.

If a scale is attached to either of the tubes, and each half-centimetre is marked a centimetre, then the reading at once gives the height of the barometer. Since, however, the bore of a glass tube is never quite uniform, two scales are fixed, one alongside each limb, having their zeros on the same horizontal plane, that alongside the closed limb reading upwards, and that alongside the open limb reading

downwards. The sum of the readings corresponding to the two mercury surfaces then gives the height of the mercury column.

In the cistern barometer the tube is straight, the open end dipping below the surface of some mercury contained in a fairly wide vessel. Since, as the atmospheric pressure alters, and therefore the height of the mercury column alters, mercury either enters or leaves the tube, the level of the mercury in the cistern will alter. As it would be inconvenient to have a cistern with such a large cross section, in proportion to that of the tube, that such fluctuations in the quantity of mercury contained in the tube as occur in practice should not *appreciably* alter the level of the surface in the cistern, a device due to Fortin is employed, by means of

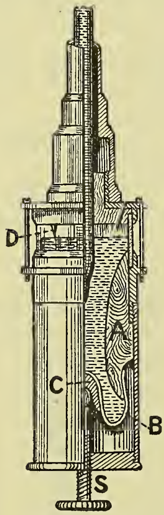


FIG. 55.

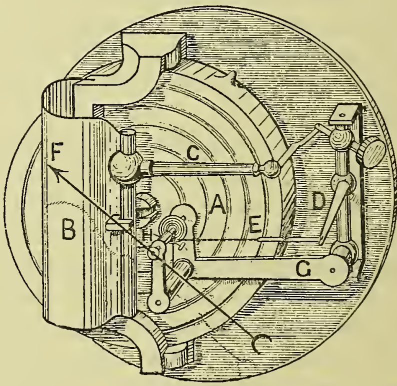


FIG. 56.

which the level of the surface of the mercury in the cistern is always brought back to a fixed mark connected to the scale by which the height of the column is measured. The plan adopted is shown in Fig. 55, and consists in making the bottom of the cistern flexible. The upper part of the cistern is of glass, and is cemented to a boxwood ring A, to which is tied a ring of buckskin B. This buckskin carries a wooden button C, which rests on the point of a screw S, and forms a flexible bottom to the cylinder, so that the surface of the mercury in the cistern can be raised or lowered by turning the screw. A small, pointed ivory pin, D, is fixed to the top of the cistern, and the graduations of the scale, which are usually engraved on a metal tube surrounding the glass barometer tube, count from the point of this pin. Before making a reading of the meniscus

of the mercury in the tube, the surface of the mercury is adjusted till it exactly touches the point of the ivory pin. This adjustment is complete when the point of the pin appears just to touch its image, as seen by reflection in the surface of the mercury.

The aneroid barometer consists essentially of a cylindrical metal box A (Fig. 56), the lid of which consists of a thin corrugated metal plate. The inside of this box is exhausted by means of a pump, leaving a more or less perfect vacuum, and the pressure of the air, acting on the thin elastic lid, bends it and forces it in to a certain extent. As the pressure of the atmosphere varies, the amount of flexure of the lid varies, and by means of a system of delicate levers, C, D, E, this change in the flexure of the lid is shown by the movement of a pointer, F, over a graduated scale. The great advantage of an aneroid barometer over a mercury barometer is its extreme portability. The scale of all aneroids, however, has to be divided by comparing their reading with a mercury barometer. It is also found that if the pressure changes rapidly the reading of an aneroid does not change quite as rapidly as the pressure.

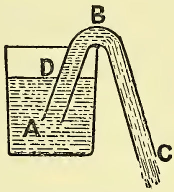


FIG. 57.

The corrections which have to be applied to the readings of a mercury barometer to allow for the effect of the temperature will be considered later (§ 61).

The syphon.

A contrivance which depends on the atmospheric pressure for its action is the syphon. The syphon consists of a bent tube ABC (Fig. 57), open at both ends, one leg being of greater length than the other. If the tube is filled with a liquid, and the end of the shorter limb dipped beneath the surface of some of the liquid, then the pressure at the end A, tending to force the liquid into the tube, is equal to the atmospheric pressure minus the weight of the column of liquid DB. The pressure at C, tending to force the liquid up the tube, is the atmospheric pressure less the weight of the column of liquid CB. Hence, since CB is greater than DB, the pressure tending to force the liquid in at A is greater than that at C, so that the liquid will flow in at A and out at C as long as the surface of the liquid is above the end A. The syphon depending, as it does, on the atmospheric pressure to force the liquid up from D to B, will not work in a vacuum, or if the height DB is greater than that of the barometric column of the given liquid. In either of these cases the liquid would not fill the bend of the syphon, but there would be a vacuum there.

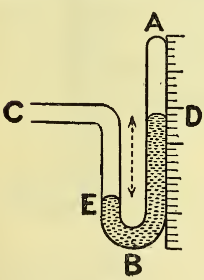


FIG. 58.

42. The Air Manometer.—The elasticity of a gas can be made use of to measure pressures. An instrument for this purpose consists of a curved tube ABC (Fig. 58) closed at one end, A, with some mercury in the bend enclosing some air in the closed limb, the volume of which can be read off on a scale attached to the side of the tube. The open end C being connected with the vessel in which the pressure has to be measured, suppose the volume of the air to be reduced from V_0, at atmospheric pressure, to V_1, the mercury in the tube standing at E and D in the two branches of the tube. Then the pressure acting through C is balanced by the elasticity of the air, together with the weight of a mercury column of height DE. The pressure due to the elasticity of the air is by Boyle's law equal to $\frac{V_0}{V_1}$ atmospheres, and hence the pressure to be measured is equal to $\frac{V_0}{V_1}$ atmospheres together with the weight of the column of mercury DE. Of course for high pressures a correction would have to be applied, to allow for the deviation of air from Boyle's law.

McLeod gauge.

For measuring very small pressures, such as exist in incandescent electric lamps, what is called a McLeod gauge is used. This gauge depends for its action on an application of Boyle's law, and consists of a glass vessel A (Fig. 59), to the top of which is attached a narrow-bore tube DC, closed at the top. A tube G attached to the lower end of A has a side branch which is connected by a bent tube E to the vessel F, the pressure in which is to be measured. To the lower end of G is attached a rubber tube which communicates with a mercury reservoir H. The reservoir H is first lowered so that the mercury surface in G is below the side tube, and hence the pressure in A is the same as in F. The reservoir is then raised, and the mercury as it rises first cuts off communication with F and then compresses the air in A into the tube CD. Let the volume of A and the tube CD be V, and that of the tube CD be v, while when the mercury has risen to C in one tube it has risen to E in the other, let the vertical distance between C and E be h. Then if p is the pressure in the vessel F, the pressure when the gas is reduced to the volume v is $p+h$. Hence by Boyle's law we have

$$pV=(p+h)v$$

or

$$p=\frac{hv}{V-v} \quad . \quad . \quad . \quad . \quad (27)$$

and thus knowing $V-v$ and v, which are obtained by weighing the mercury required to fill the vessel A and the tube CD, we can calculate the pressure in F, and since v is very small compared to $V-v$, h is much larger than p, and hence can be measured with considerable accuracy.

43. Air-pumps.—An air-pump is an instrument for withdrawing the air from within a vessel. In its simplest form the air-pump consists of a cylinder in which a piston P (Fig. 60) fits air-tight. There is a hole through the piston closed by a flap valve C, which can open outwards. A pipe, the opening to which can be closed by a valve, B, which opens inwards, leads to the vessel D, that is to be exhausted. When the piston is drawn upwards the valve C closes, and the pressure below the piston is reduced so that the air in the receiver, on account of its elasticity, is able to raise the valve B, and flows into the cylinder. When the piston descends the valve B closes, and the air in the cylinder is compressed till it is able to force open the valve C, and escape into the air. By repeating this process the air is gradually pumped out of D.

Let the volume of the vessel D and the pipe connecting it to the cylinder

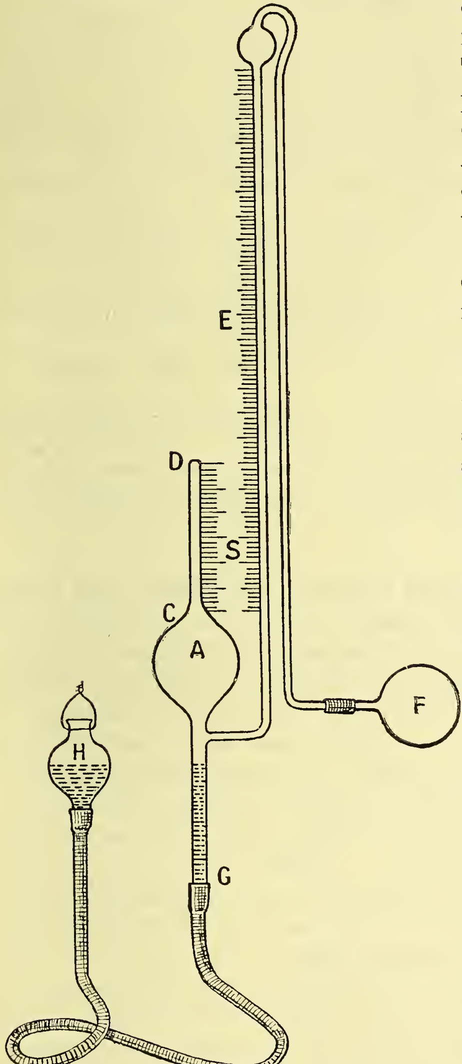


FIG. 59.

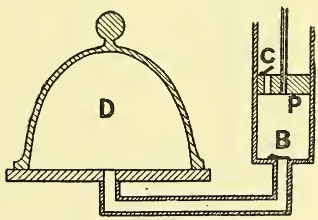


FIG. 60.

be V, and the volume of that part of the cylinder through which the lower surface of the piston moves during a stroke be v. Then, if we start with the piston at the bottom of its stroke, the volume of the mass (m) of air in the instrument is V. At the end of the upward stroke the

volume of this mass of air will be $V+v$. Of this volume v c.c. will be expelled at the down-stroke, and V c.c. will be left in the instrument. Hence at the end of the first stroke the mass of air in the receiver is $\frac{V}{V+v}m$. At the end of the second up-stroke the volume of this mass of air expands to $V+v$, and during the down-stroke v c.c. of air at this density are expelled. Hence the mass of air at the end of the second stroke left in the receiver is $\frac{V}{V+v}$ times the mass of air in the receiver before the second stroke, or $\frac{V}{V+v} \cdot \frac{V}{V+v} m$, which may be written $\left(\frac{V}{V+v}\right)^2 m$. In the same way the mass of air left after three strokes is $\left(\frac{V}{V+v}\right)^3 m$, and generally the mass of air left after n strokes is $\left(\frac{V}{V+v}\right)^n m$. Since the pressure after n strokes will be to the initial pressure, p, as the mass of gas contained in the receiver after n strokes is to the mass contained at the start, we have that the pressure after n strokes is $\left(\frac{V}{V+v}\right)^n p$.

With such a pump as described above it is not possible to get a very low pressure, owing to the fact that the piston when at the bottom of the stroke does not completely expel all the air in the cylinder, since there is always a small space between the bottom of the piston and the top of the valve B. As exhaustion proceeds a time will come when the cylinder—full of gas at the pressure of the receiver—will at atmospheric pressure only just fill the clearance below the piston, and hence no air will be expelled through the valve C. Another defect is that at low pressures the elasticity of the gas is insufficient to lift the valve B, so that the gas does not flow from the receiver into the pump-barrel.

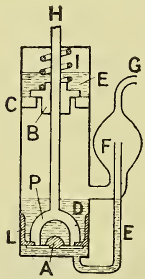


FIG. 61.

In the Fleuss pump, which is shown diagrammatically in Fig. 61, the above defects of the simple air-pump are avoided. The piston P consists

Fleuss oil-pump.

of a metal frame with a leather bucket L, and has a valve A, which opens upwards. The pressure of the air on the upper side of the piston presses the leather against the wall of the cylinder, and thus ensures a close fit. The valve A only acts when exhaustion commences, and also to allow any oil which may have got below the piston to pass up. The piston-rod passes through the upper valve B, which is held down on the seating C by means of a

spring I. The communication with the vessel to be exhausted is made through the tube G. The tube E is designed to relieve the piston during the first few strokes, when otherwise there would be a vacuum below and atmospheric pressure above. As the piston rises it cuts off communication with G, and then compresses the air till it strikes the valve B, which it raises, allowing the air to escape. The whole of the air is driven out, some of the oil D, which is above the piston, being driven out. The valve does not close till the piston has descended about $\frac{1}{4}$ inch, so that some of the oil above the valve passes down to the top of the piston.

A very good mechanical pump will exhaust a vessel till the pressure of the remaining air will support a column of mercury of about 0·05 millimetre in height. In order to get a better vacuum than this, it is necessary to employ a pump in which the piston is formed by a quantity of mercury. One form of such mercury pumps is shown in Fig. 62, and is called a Töpler pump. A cylindrical glass vessel A has a side-tube B attached, and to the lower end of this side-tube is attached another tube CH, which is connected to the vessel to be exhausted. To the top of A a tube DG, about 80 cm. long, is attached, while to the bottom another tube EF, also about 80 cm. long, is attached. The lower end of EF is connected by a length of rubber tubing with a reservoir K containing mercury. When K is raised the mercury rises in FE, and when the surface reaches C it cuts off the connection between the vessel A and the tube CH. K is raised till the mercury completely fills A and flows out through G, driving any air that was in A before it. If now K is lowered, so that the surface of the mercury in K is more than 76 cm. below C, the mercury will fall in A and in DG till it stands at the barometric height in DG, and will leave a Torricellian vacuum in A. When the mercury in EF has fallen below C, the tube HC will be connected to this vacuum, and hence the air in the tube HC and any vessel attached to H will rush into A. By again raising K the air enclosed in A will first be cut off from CH by the rising mercury and then forced out of the apparatus at G, and on lowering K a vacuum will again be left in A. The mercury here plays the part of a piston moving up and down in the cylinder A, and constitutes its own valves. By repeating this operation a number of times it is possible to obtain a

Töpler mercury pump.

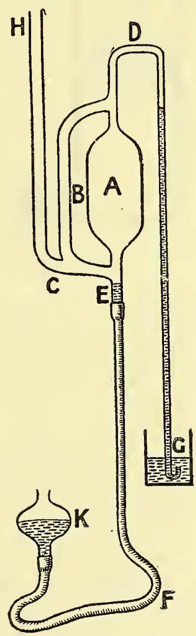


Fig. 62.

very good vacuum, in fact the pressure inside a vessel has been reduced to 0·000012 millimetre of mercury, that is, to ·000000016 or ·016 of a millionth of an atmosphere.

Compression pumps.

Pumps are also used to compress air, that is, to force it into a vessel in which the pressure is higher than atmospheric. A familiar example of such a pump is that used to blow up the tyres of a cycle or motor-car. A section of such a pump is shown in Fig. 63. The piston P consists of a cup-shaped piece of leather clamped between two metal discs. On the upward stroke the edge of the leather leaves the side of the barrel, allowing the air to pass from above to below the piston. On the downward stroke, the pressure of the air confined below the piston forces the leather against the wall of the barrel, and so prevents the air passing the piston. A valve A allows the air to flow into the vessel, which is connected to the pipe B on the downward stroke, but closes on the upward stroke. The valve A is in some pumps placed close to the barrel, while in other pumps, particularly cycle pumps, it is at the far end of the tube B, that is, at the tyre itself. There is a very marked difference in the efficiency of the pump according to the position of this valve. Suppose that the volume swept out by the piston during the stroke is V, while the volume enclosed between the bottom of the piston and the valve A when the piston is at the bottom of the stroke is v. Further, let the pressure in the tyre be P atmospheres.

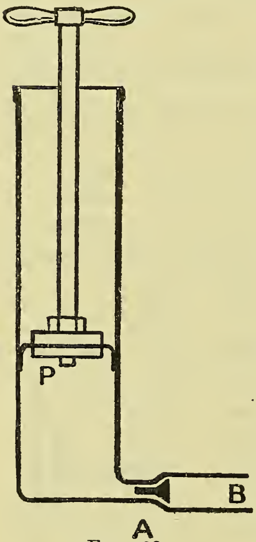


FIG. 63.

The volume of the air between the piston and the valve A at the top of the stroke is $V+v$, and its pressure is at 1 atmosphere. As the piston descends the pressure increases to P, and then air starts flowing past the valve A into the tyre. At the bottom of the stroke we have left in the pump a volume v of air at a pressure P. Now this quantity of air occupied at atmospheric pressure a volume vP, if we suppose that the temperature has remained constant during the compression. Hence the air which has entered the tyre would at atmospheric pressure occupy a volume $V+v-vP$ or $V-(P-1)v$. Thus the air sent into the tyre is less than the volume swept out by the piston by an amount $(P-1)v$, that is the product of the difference between the pressure inside and outside the tyre into the volume between the piston at the bottom of the stroke and the valve A. Hence for comparatively high pressures, such as are used in a motor tyre, it is most important that v should be kept

small. Thus suppose V is 10 cubic inches, v 1 cubic inch, and the pressure in the tyre is 80 lbs. per square inch *above* atmospheric, *i.e.* 95 lbs. per square inch absolute. Then the volume swept out by the piston is 10 cubic inches, and of this only 4·6 cubic inches is forced into the tyre, the rest remaining in the pump.

CHAPTER IX

PROPERTIES OF LIQUIDS

44. Equilibrium of a Liquid at Rest.—In the case of a liquid at rest under the influence of gravity the free surface must be horizontal. If it were inclined to the horizon, then the weight of a particle P (Fig. 64) of the liquid at the surface would have a component parallel to the surface of the liquid. Since the surface is everywhere at the same pressure, there would be nothing in the nature of a hydrostatic pressure to resist this force, and as the liquid itself would offer no resistance, the particle P would move along the surface, and hence the liquid would not be at rest.

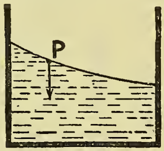


FIG. 64.

Although a comparatively small surface of a liquid is for all practical purposes plane, it is not absolutely so, and when dealing with large surfaces, this departure from planeness becomes appreciable. The condition that the particle P (Fig. 64) should be at rest is that the line of action in the attraction of gravity on P should be normal to the surface at P. Hence the surface of a liquid is always normal to the radius of the earth at the point considered, and therefore forms part of a sphere having the earth's radius as radius. The fact that the surface of a liquid is always horizontal is made use of in the spirit-level. This consists essentially of a glass tube ABC (Fig. 65), which is slightly bent, and fitted, with the convex surface upwards, in a frame DE. This tube is closed at either end, and is filled with alcohol[1] except for a bubble of air B, which is left in. This bubble constitutes the only free surface of the liquid, and it always sets itself at the highest point of the curved tube. Hence, if the tube is fixed in the frame in such a way that when the lower surface of the frame is horizontal the highest point of the bent tube is opposite a fixed mark on the top of the tube, then, whenever the bubble is opposite this mark, the lower surface of the stand will be horizontal. If one end, say E, is tilted up, then the marked point of the tube is no longer the highest, and the bubble moves towards E.

The spirit-level.

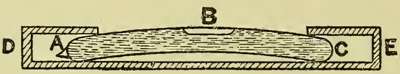


FIG. 65.

[1] Alcohol is used on account of its extreme mobility.

Suppose a U-tube ABC (Fig. 66) has the same liquid in either limb, then the two surfaces A and B will be in the same horizontal plane. For if we consider a point D within the liquid, at a depth h_1 below the surface at A, and at a depth h_2 below the surface at C, then the pressure at D must be the same, whether caused by the column AD or the column CD; otherwise the liquid would move towards the side on which the pressure was least. Hence (see § 45)

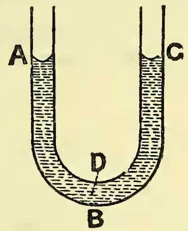


FIG. 66.

$$h_1 gD = h_2 gD,$$

where D is the density of the liquid.

$$\therefore \qquad h_1 = h_2.$$

By an exactly similar line of argument it can be shown that the pressure at any pair of points, one in either limb, is the same if these points lie in the same horizontal plane.

Balancing columns of liquid.

If, instead of having the same liquid in both limbs, one limb AB (Fig. 67) contains a liquid of less density than that in the other; then, if B is the surface separating the two liquids, from what has been said in the previous paragraph, the pressure at a point D in the denser liquid in the same horizontal plane as B must be equal to the pressure at B. Hence the pressure exerted by the column AB of the one liquid must be equal to the pressure exerted by the column CD of the other liquid. So that, if h_1 and h_2 are the heights of these columns, and D_1 and D_2 are the densities of the liquids, we have—

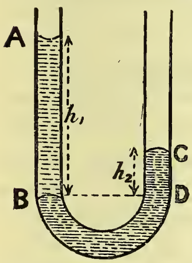


FIG. 67.

$$h_1 D_1 g = h_2 D_2 g,$$

or

$$\frac{h_1}{h_2} = \frac{D_2}{D_1} \quad . \quad . \quad . \quad (28)$$

That is, the heights of the columns of the two liquids above the level of their common surface are to one another inversely as the densities of the liquids.

45. Pressure on a Surface exposed to a Liquid. Centre of Pressure.—The difference in pressure between a point at a depth h below the surface of a liquid of density D and the pressure at the surface is gDh. For consider a prism of the liquid of unit cross-section extending vertically from the surface to the depth h, the pressure on the sides have a resultant zero, and if p is the pressure acting on the top and P that acting on the bottom, the resulting upward pressure is $P-p$, and this by the principle of Archimedes must be equal to the weight of the cylinder of fluid. The volume

Variation of pressure with depth.

of the cylinder is $h \times 1$, and its weight measured in absolute units is gDh, so that

$$P - p = gDh \quad . \quad . \quad . \quad . \quad (29)$$

If h is measured in centimetres, g is about 980, and the pressure is obtained in dynes per square centimetre. If we omit the g the pressure Dh is given in grams weight per square centimetre.

If s is the specific gravity of the liquid and h is measured in feet, then the pressure at a depth h is given by

$$62 \cdot 42 \, sh \text{ lbs. per square foot,}$$

or

$$\frac{62 \cdot 42}{144} sh \text{ lbs. per square inch.}$$

Pressure on a vertical surface immersed in a liquid.

We have seen in § 39 that the pressure exerted by a fluid at any point on a surface is always perpendicular to the surface, and we now proceed to calculate what will be the total pressure on a surface exposed to a liquid, say on the side of a box filled with a liquid. Let ABCD (Fig. 68) be the side of such a box, the surface of the liquid being at AB. Take G, the middle point of the face, and suppose the face divided into an even number of very narrow horizontal strips each of width y. Consider two of these strips, EF and HJ, which are at the same distance X on opposite sides of G. If X is the depth of G below the surface, the pressure[1] at the strip EF is $D(X-h)$ and that at HJ is $D(X+h)$ where D is the density of the liquid. Hence, since the area of each strip is $DC \times y$, the sum of the total pressures exerted in the *two* strips is $D(X-h)y.DC + D(X+h)y.DC$, or $D.X.2y.DC$. But $D.X$ is the pressure at G, and $2y.DC$ is the sum of the areas of the strips. Hence we see that the total pressure exerted by the fluid on these two strips is the same as would be exerted on a strip having an area equal to the sum of their areas if the pressure were uniform and equal to that at G. Proceeding in the same way for all the other pairs of strips, we get that the total pressure on ABCD is equal to the area of the face multiplied by the pressure at G. We may now proceed to apply the same method to the case of a vertical surface of any shape.

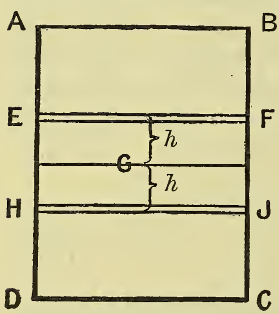


FIG. 68.

[1] In order to avoid having to introduce the symbol g, we will suppose the pressures to be measured in gravitational units, *i.e.* grams weight per square centimetre, or pounds per square inch. Further, we shall not consider the pressure acting on the *surface* of the liquid, since this will only be a constant quantity added to the number we find. In the case considered this atmospheric pressure would in general act on the outside of the box, and hence the resultant outwards pressure is that due to the weight of the liquid only.

Let AB (Fig. 69) be the surface of the liquid and IEFJ any plane vertical surface immersed in the liquid, G being the centre of gravity of the surface at the depth X. Divide the surface into two parts by the horizontal line KL through G. If we take a narrow horizontal strip EF of area a_1 at a height h_1 above G, the pressure at this strip is $\mathrm{D}(\mathrm{X}-h_1)$ and the total pressure on the strip is $\mathrm{D}(\mathrm{X}-h)a_1$. Proceeding in this way for all the strips above KL the total pressure on this part of the surface is

$$DX(a_1+a_2+a_3,\ \&c.)-(h_1a_1+h_2a_2+h_3a_3,\ \&c.)D.$$

Similarly for the lower portion of the surface the total pressure is

$$DX(a'_1+a'_2+a'_3,\ \&c.)+(h'_1a'_1+h'_2a'_2+h'_3a'_3,\ \&c.)D.$$

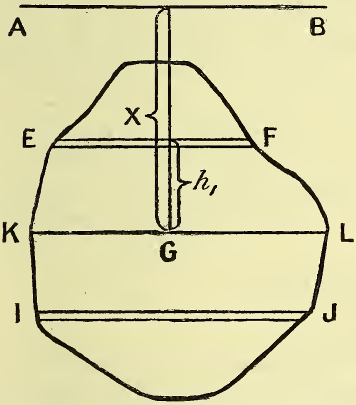


FIG. 69.

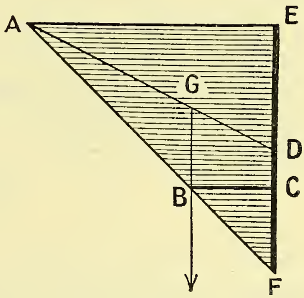


FIG. 70.

Now since G is the centre of gravity of the surface by the principle of moments as used in § 22, we have

$$h_1a_1+h_2a_2+h_3a_3+\&c.=h'_1a'_1+h'_2a'_2-h'_3a'_3,\ \&c.$$

Further,

$$a_1+a_2+a_3,\ \&c.+a'_1+a'_2+a'_3+\&c.=\text{total area of surface}=A\ \text{(say)}.$$

Hence Total pressure on surface $=DXA$ . . (30)

Thus the total pressure is equal to the area of the surface multiplied by the pressure which exists at the centre of gravity.

The above result also applies to the case of an inclined plane surface, but this extension may be left to the reader.

Having found the magnitude of the total pressure acting on the rectangular vertical surface shown in Fig. 68, that is, the magnitude of the

resultant pressure, we now require to find the point of the surface at which it may be supposed to act. In the first place it is evident from symmetry that it must lie somewhere on the vertical through the mid point of the surface.

Centre of pressure.

Suppose that we had an inclined surface AF and a vertical surface EF (Fig. 70) in the fluid, and that we consider the forces acting on the slice of fluid taken parallel to the paper and of unit thickness. The portion of fluid AEF has its centre of gravity at G on the line joining A to the mid point of EF, and such that DG is a third of AD. The weight of the fluid acts vertically through G, and hence since there is equilibrium the resultant reaction due to the walls must act through G. Now the pressure on EF is everywhere horizontal, so that the reaction is also horizontal and cannot therefore have any *vertical* component. Hence it follows that the *vertical* component of the reaction due to AF must act through G, that is, must pass through B. In other words, the resultant pressure in the plane AF must pass through B. Next, since the horizontal component of the reaction of AF must be equal to the reaction due to EF, or the liquid prism would not be in equilibrium, since gravity has no horizontal component, it follows that the reaction of EF must act through the point C. Thus the resultant pressure of the liquid must pass through the point C, and this point is called the centre of pressure. From similar triangles we have

$$\frac{FC}{FE}=\frac{FB}{FA}=\frac{DG}{DA}=\frac{1}{3}.$$

Hence the centre of pressure of a rectangle which has one edge at the surface of the fluid is at a depth equal to two-thirds of the depth of the lowest edge below the surface. Further, this result holds whether the surface is vertical or inclined.

The determination of the centre of pressure in more complicated cases would be beyond the scope of this book, and hence the student wishing to pursue the subject further must refer to text-books on hydrostatics.

46. Flotation. Hydrometer.—Since when a body is immersed in a fluid it experiences an upward force, due to the pressure of the fluid, equal to the weight of the fluid displaced, it follows that if the density of the body is less than that of the fluid, the weight of the displaced fluid will be greater than the weight of the body, and hence the upward force acting on the body due to the fluid will be greater than the downward force due to gravity. If no other forces are acting on the body, it will therefore rise until the weight of the displaced fluid is exactly equal to that of the body. In the case of a body such as a balloon in the air, this will happen when it has risen to such a distance that the density of the air has become so much reduced that the weight of air displaced by the balloon

is equal to its own weight. In the case of a solid immersed in a liquid, the body will rise till, some of the solid having risen above the surface of the liquid, the weight of the volume of liquid displaced by the remainder, which is still submerged, is equal to the weight of the body.

In order that a body floating in a liquid may be in equilibrium, not only must the upward pressure due to the liquid be equal in magnitude to the weight of the body, but it must also act vertically upwards through the centre of gravity of the body. If we suppose the body removed and replaced by some of the liquid which has become solid and occupies exactly the position of the *immersed* part of the solid, this solidified portion of liquid will be in equilibrium. Hence, since its weight acts vertically through its centre of gravity, the pressure due to the part of the liquid which has remained liquid must also act vertically through the centre of gravity of the solidified portion. The direction and magnitude of this pressure must be the same as that which was acting on the floating body, so that we see that the upward force due to the liquid is equal to the weight of the liquid displaced, and acts in a vertical direction through the point which would be the centre of gravity of the *displaced liquid*. If G (Fig. 71, A) is the centre of gravity of a floating body, and H the centre of gravity of the displaced water, the two points G and H must, if the body is in equilibrium, be vertically one over the other. If the body be displaced into some such position as that shown at B, then the centre of gravity of the displaced liquid will no longer be at H, but at some point such as H′. The body is then acted upon by a couple which tends to bring it back into the position shown at A. The point M, where the new vertical through H′ cuts the vertical drawn through H in the undisturbed position, is called the *metacentre*. In the above figure the metacentre is above the centre of gravity, and the floating body is in stable equilibrium, as the couple when it is deflected tends to restore it to its original position. In the case shown in Fig. 72, however, the couple, which comes into play when the body is deflected from the A position, tends to increase the displacement, and

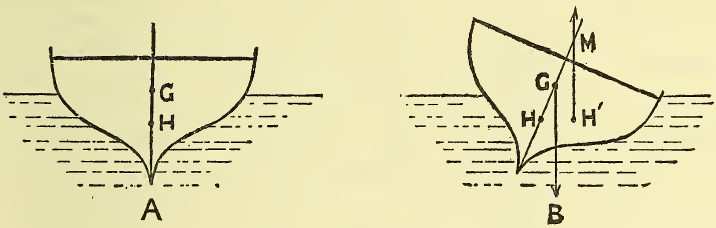


FIG. 71.

Metacentre.

hence the condition figured at A is unstable. In this case it will be seen that the metacentre M is *below* the centre of gravity of the floating body.

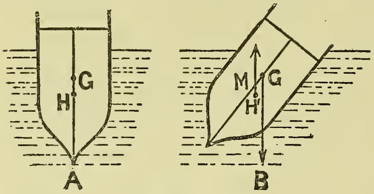


FIG. 72.

Hence a floating body is stable when the metacentre is above the centre of gravity, and the higher the metacentre is above the centre of gravity, the more stable is the body. If the metacentre coincides with the centre of gravity of the floating body, as it does in the case of a sphere, the body is in neutral equilibrium, while if the metacentre is below the centre of gravity the equilibrium is unstable. These principles are of great importance in designing ships, the object of ballast being to lower the centre of gravity so as to keep it well below the metacentre.

The amount of a floating body immersed in a liquid depends on the density of the liquid, for the product of the density of the liquid into the volume of the immersed part, which gives the weight of liquid displaced, must always be equal to the weight of the body.

Hydrometer.

Hence the volumes of that part of a floating body immersed in different liquids may be employed to compare the densities of liquids. An instrument depending on this principle, called the hydrometer, is shown in Fig. 73. It consists of a glass bulb B, to the lower end of which a small bulb A is fixed, and at the upper end a narrow glass stem CD. Some mercury or lead-shot is placed in A, so that the instrument floats upright. The stem CD is graduated, so that the depth to which the instrument sinks in the different liquids can be read off. The graduations are generally arranged so that the reading on the stem gives the density of the liquid directly.

Another form of hydrometer is shown in Fig. 74, and is called Nicholson's hydrometer. In this instrument the stem is not graduated, but has a single mark, O, and when in use the instrument is always sunk to this mark, so that the volume immersed is constant. Attached to the float A are two scale-pans B and C, the lower one being weighted so that the instrument can float upright.

When using this instrument to determine the density of a liquid, it is first floated in water at a known temperature, and weights are placed in the upper scale-pan till the mark O is coincident with the surface of the liquid. Let W be the weight of the instrument itself, and w_1 the weights added; then the weight of the water displaced is $W+w_1$, and the volume V of the displaced water is given by

$$V=(W+w_1)/\Delta,$$

where Δ is the density of the water at the given temperature.

Now let the instrument be floated in a liquid of density D, and let the weight which has to be placed in the pan B to bring the mark to the surface of the liquid be w_2. Then the weight of the liquid displaced is $W+w_2$. Now, since the volume of the hydrometer immersed is the same as before, namely V, we have—

$$D=(W+w_2)/V$$
$$=\frac{W+w_2}{W+w_1}\cdot\Delta \quad . \quad . \quad . \quad . \quad (31)$$

This instrument is more often used for finding the density of solids than of liquids. For this purpose the hydrometer is floated in water, and the solid placed in the pan B, and weights w_3 added till it sinks into the sighted position. Since, when the solid is not present, the weight necessary to sink the instrument is w_1, it follows that the weight of the solid is w_1-w_3. Next, the solid is placed in the lower pan C, and the weight w_4 necessary to sink the instrument determined. The solid being immersed in water, will lose in weight an amount equal to the product $V\Delta$, where V is its volume. Hence

$$V\Delta=w_4-w_3$$

or
$$V=(w_4-w_3)/\Delta$$

Therefore the density of the solid is given by

$$D=\frac{w_1-w_3}{w_4-w_3}\Delta \quad . \quad . \quad (32)$$

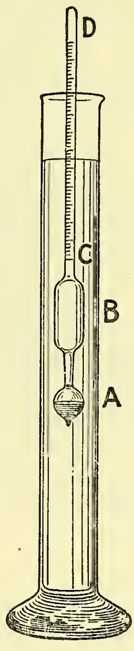


Fig. 73.

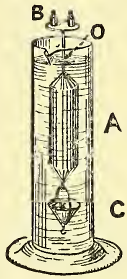


Fig. 74.

47. Density of Liquids.—In order to determine the density of a liquid, the mass of a known volume must be measured. If, however, we know the density of any given liquid, then we can find the density of any other liquid by comparing the mass of any volume of the liquid with the mass of an *equal* volume of the standard liquid. The liquid almost always used as a standard is water. Since, however, the density of water is different at different temperatures, as is shown by the curve given in Fig. 73, the density corresponding to the temperature at which the water was measured has to be employed in the calculations.

Change of density of water with temperature.

It will be noticed (Fig. 75) that the density of water is a maximum and equal to unity at a temperature of 4° C. Since it is difficult to always adjust the temperature to 4°, it is more usual to arrange that the water has a temperature of 0°, for this temperature can easily be secured by means of melting ice.

One method of comparing the density of a liquid with that of water is to determine the loss of weight of a solid, which is unacted upon by either liquid, when weighed first in water and then in the liquid. In this way the weights or masses of equal volumes of the liquid and of

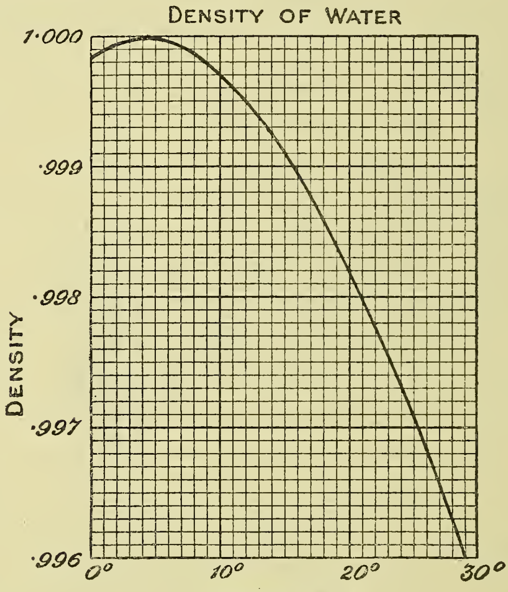


FIG. 75.

water are obtained. If m_1 is the loss of weight in the given liquid of density D, and m_2 is the loss of weight in water of density Δ, then

$$m_1 = VD$$

and

$$m_2 = V\Delta,$$

where V is the volume of the solid. Hence

$$D = \frac{m_1}{m_2}\Delta \quad . \quad . \quad . \quad . \quad . \quad (33)$$

and, by taking the value of Δ for the temperature of the experiment from the curve given above, D can be calculated.

Another method in common use for determining the density of a liquid is to weigh a small bottle, called a specific gravity bottle or pyknometer, when full, first of water, then of the liquid. Two forms of pyknometer which are in common use are shown in Fig. 76. One consists of a small glass bottle A, fitted with a well-ground-in glass stopper B. This stopper is pierced by a fine hole. The bottle is completely filled with the liquid, and the stopper inserted, care being taken not to include any air-bubbles. The superfluous liquid flows out through the hole in the stopper and is wiped off. The other form consists of a bent glass tube CDE of the shape shown. The end C is drawn out into a fine capillary, and a fine mark is engraved on the glass at F. To fill the tube the end C is dipped below the surface of the liquid, which is drawn into the tube by suction at E till it fills it a little above the mark F. Then, by touching the capillary C with a piece of blotting-paper, liquid is withdrawn till the surface comes exactly to the mark F.

The specific gravity bottle.

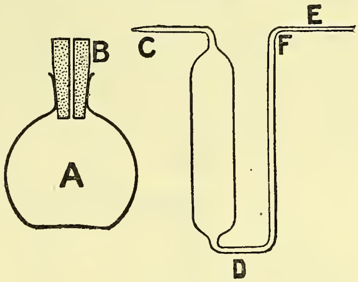


FIG. 76.

Let w_1 be the weight of the empty pyknometer, and w_2 and w_3 the weight when full of the liquid and water respectively. Then $w_3 - w_1$ is the weight of water which fills the instrument. Hence if Δ is the density of the water at the temperature at which the pyknometer was filled, its volume V is given by

$$V = (w_3 - w_1)/\Delta.$$

The weight of a volume V of the given liquid is $w_2 - w_1$. Hence the density D of the liquid is

$$D = \frac{w_2 - w_1}{w_3 - w_1} . \Delta \qquad (34)$$

48. Elasticity of Liquids.—Liquids exhibit bulk elasticity alone, that is, they can only offer *permanent* resistance to changes in volume. The change in volume produced by change in pressure is very small in the case of liquids, hence their elasticity is very great. Thus an increase in pressure of 1 metric atmosphere produces a decrease of ·000047 c.c. in 1 cubic centimetre of water at a temperature of 8°. The elasticity of water is therefore $\frac{10^6}{\cdot 000047}$ or $2{\cdot}1 \times 10^{10}$ dynes per square centimetre. Comparing this with the elasticity of a gas at 1 metric atmosphere, namely 10^6 dynes per square centimetre, it will be seen how very much greater is the elasticity of a liquid than that of a gas. For most practical purposes liquids may be taken as incompressible.

49. Hydraulic Press.—Pascal's law, that liquids transmit in all directions and without diminution any pressure that is applied to them, receives an important application in the hydraulic press or ram. This machine was invented in 1795 by Bramah, and is shown in section in Fig. 77. It consists of a large metal cylinder, A, with very strong sides, in which an iron ram works water-tight, through a gland, B. This gland is made water-tight by means of a circular leather washer, the section of which is U-shaped, the concave surface being turned towards the inside of the cylinder. The pressure of the water in the cylinder forces this washer against the ram on the inside, and against the neck of the cylinder on the outside, so that the greater the pressure of the water the more tightly does the washer fit. The cylinder is connected by a strong pipe, C, with a force-pump, of which the piston, D, is of small diameter. By

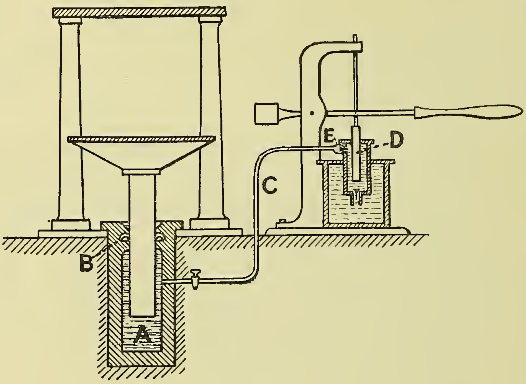


FIG. 77.

this means water can be pumped into the large cylinder A. When the plunger of the pump is forced down, the liquid in the machine transmits the pressure to the base of the ram, which is forced up. If a is the area of cross-section of the plunger of the pump, and the downward force exerted on the plunger is P, then the pressure exerted on the water in the pump is P/a. This pressure is transmitted to the cylinder A, and hence a pressure of P/a acts on each unit of surface of the base of the ram. If A is the area of cross-section of the ram, the total upward force exerted on it is AP/a. In other words, the force (W) exerted by the ram is to the force acting on the plunger of the pump as the area of cross-section of the ram is to that of the plunger, or

$$\frac{W}{P}=\frac{A}{a}=\frac{D^2}{d^2} \quad . \quad . \quad . \quad . \quad (35)$$

if D is the diameter of the ram and d the diameter of the plunger.

The principle of the hydraulic press is also employed as a means of storing power, which is required in an intermittent manner, such as for working lifts. In this case powerful pumps are employed to pump water into a strong steel receiver fitted with a wide piston, like the ram of a Bramah press, which is loaded with heavy weights. The work performed by the engine which drives the pumps is employed in raising these weights, and the potential energy thus stored up can be usefully employed by connecting the receiver by pipes to the hydraulic engines to be driven by the water under high pressure.

50. Liquids in Motion. Bernouilli's theorem.—Let AC (Fig. 78) be a vessel filled with a liquid of density D, so that unit volume of the liquid, since it has a mass D, will be attracted by the earth with a force gD in absolute units. If the bottom of the vessel is at the level of the ground, then the potential energy of unit volume of the liquid at A is gDH, where H is the height of A above the bottom. Let us now slowly draw off the liquid through a tube E, and at the same time let liquid flow in at the top so as to keep the level of the surface of the liquid the same. The unit volume of liquid we have been considering will gradually sink, and when it reaches the point B at a height h above the bottom its potential energy due to its *height* will be gDh. If the liquid has been drawn off very slowly, no appreciable kinetic energy will have been acquired. Further, no appreciable work will have been done against friction, for as we shall see later, friction in liquids is proportional to the velocity between different parts of the liquid, and hence if the velocity is very small the friction can be made negligible. Hence the unit volume of the liquid has lost potential energy of amount $gD(H-h)$, but since no kinetic energy has been acquired or any energy wasted we conclude that when at B the liquid must possess potential energy in addition to that due to its height above the ground. Now between A and B the pressure to which the liquid is subjected has increased by an amount $gD(H-h)$, and this is equal to the missing potential energy. Hence we conclude that the potential energy of the unit volume of the fluid consists of two parts, one due to the height above the ground and the other due to the pressure to which it is subjected, so that the total potential energy of the unit of volume at B is $gDh+p$, where p is the difference in the pressure at the point B and a point at the surface of the earth and outside the liquid. Another way of deducing the above is to consider that to force a unit volume of liquid in at the level of B we should have to do an amount of work p. For suppose we used a pump

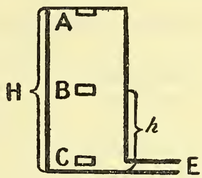


Fig. 78.

Potential energy of a liquid due to the pressure.

of which the cross-section was unity, then the force with which the piston would have to be driven in is p, while the movement of the piston necessary to introduce unit volume would be unity, and hence the work done is $p \times 1$.

If when at B the unit of volume has a velocity v, then on this account its kinetic energy is $\frac{1}{2}Dv^2$, and hence the total energy of unit volume would be

$$gDh + p + \tfrac{1}{2}Dv^2.$$

If no external forces, other than gravity, act on the unit of volume considered, then its total energy must remain constant, and hence

$$gDh + p + \tfrac{1}{2}Dv^2 = \text{constant}; \qquad . \qquad . \qquad (36)$$

a result which is known as Bernouilli's theorem, and is of fundamental importance in all branches of hydraulics.

Bernouilli's theorem.

As an example of the application of this theorem, let us consider the motion of water in a horizontal tube with a constricture in it, such as is shown in Fig. 79. If the cross-sections of the tube at A, B, and C are a, b, and c respectively, while the velocity with which the liquid is moving at these points is v_a, v_b, and v_c, it is at once evident that $\frac{v_a}{v_b} = \frac{b}{a}$ and $\frac{v_b}{v_c} = \frac{c}{b}$, for the same quantity of water must pass through the tube at A, B, and C in unit time, since water is practically incompressible. Since the tube is horizontal the mean level of the water at each cross-section is the same, and therefore we may take h as the same throughout. Hence, applying equation (36) we get

$$gh + p_a + \tfrac{1}{2}v_a^2 = gh + p_b + \tfrac{1}{2}v_b^2 = gh + p_c + \tfrac{1}{2}v_c^2 \qquad . \quad (37)$$

or
$$\frac{p_a}{p_b} = \left(\frac{v_b}{v_a}\right)^2 \text{ and } \frac{p_b}{p_c} = \left(\frac{v_c}{v_b}\right)^2$$

so that
$$\frac{p_a}{p_b} = \left(\frac{a}{b}\right)^2 \text{ and } \frac{p_b}{p_c} = \left(\frac{b}{c}\right)^2 \qquad . \qquad . \qquad . \quad (38)$$

Thus, since the cross-section at A is greater than the cross-section at B, the pressure at A must be greater than the pressure at B. In the same way the pressure at B is less than the pressure at C. We could have deduced this result from the consideration that since an element of volume of water is moving faster at B than at A, there must be a force tending to drive it from A towards B, and the only force which can exist is a difference in pressure, and therefore the pressure at A must be greater than that at B.

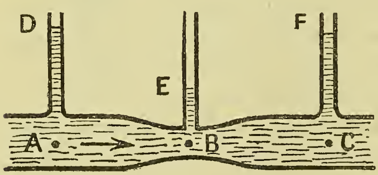


FIG. 79.

The above result is applied in the Venturi meter to measure the flow of a liquid through a pipe. If vertical pipes, D, E, and F, are fitted as shown in Fig. 77, then the difference in level between the liquid surface in D and E gives the difference in pressure between A and B. But from equation (37) above,

The Venturi water meter.

$$p_a - p_b = \frac{v_b^2}{2} - \frac{v_a^2}{2} = \frac{v_a^2}{2}\left\{\left(\frac{a}{b}\right)^2 - 1\right\}$$

or

$$v_a^2 = \frac{2(p_a - p_b)}{\left(\frac{a}{b}\right)^2 - 1}$$

But the total volume of fluid which passes the cross-section at A in unit time is av_a. Hence the volume

$$= a\sqrt{\frac{2(p_a - p_b)}{\left(\frac{a}{b}\right)^2 - 1}} = k\sqrt{(p_a - p_b)}$$

where k is a constant for any given tube. Thus by measuring the difference of pressure, $p_a - p_b$, the quantity of liquid flowing can be calculated.

Next we may apply the theorem to calculate the velocity with which the liquid will flow out of a tube such as that shown at E (Fig. 76). We suppose that the cross-section of the vessel is so great that the downward velocity of the particles of the liquid at the surface A is negligible, and hence, since at the surface $p = 0$, equation (36) gives

$$gDH = \text{constant} \quad . \quad . \quad . \quad (a)$$

Next, for unit volume which has just escaped, $p = 0$ and $H = 0$, so that by equation (36)

$$\tfrac{1}{2}Dv^2 = \text{constant} \quad . \quad . \quad . \quad (b)$$

Thus equating the left-hand sides of (a) and (b),

$$v = 2\sqrt{gH} \quad . \quad . \quad . \quad . \quad (39)$$

It must be remembered that v is the velocity of efflux of a particle of water in the issuing jet, and this result shows that this velocity is independent of the density of the liquid and the size of the hole. Equation (39) expresses what is known as *Torricelli's* law.

Torricelli's law of efflux of a liquid.

The volume of liquid which will escape per second through an opening of cross-section a is found to be considerably less than va, where v is the velocity of efflux given by Torricelli's law. This is due to the fact that the cross-section of the jet of liquid as it leaves the opening becomes contracted, forming the *vena contracta*. The cross-section of the vena contracta varies with the shape of the edge of the hole. With

a circular hole having a sharp edge the cross-section of the vena contracta is about ·62 of that of the hole.

51. Force exerted by a Jet of Fluid.—Suppose that a jet of fluid, C (Fig. 80), of cross-section a moving with a velocity v impinges on a plane surface AB, the direction of the jet being perpendicular to the surface. If D is the density of the fluid, the mass which strikes the plate in unit time is aDv, and the momentum it possesses before impact in a direction perpendicular to the plane is aDv^2. After impact the fluid moves in a direction parallel to the surface of the plate, and hence has no momentum in the original direction. Hence the change in momentum perpendicular to the plate in unit time is aDv^2, and this, therefore (Newton's second law), is the force the plate has to exert on the jet in the direction DC, while the force exerted by the jet is of equal amount, but in the direction CD. If in place of being plane, the surface is curved so that the *direction* of the movement of the water is reversed by the impact, as shown in Fig. 81, the force exerted is twice that given above, since the change of momentum is now $2aDv^2$.

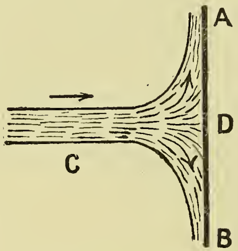


FIG. 80.

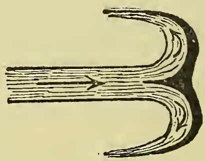


FIG. 81.

Curved buckets of this shape are used in the Pelton wheel turbine, which consists of a wheel to the edge of which are attached a number of these buckets, one or more jets of water being directed so as to impinge on the buckets.

Liquid jet impinging on inclined surface.

Next let us consider the case of a jet impinging on a plate at an angle θ (Fig. 82). Before impact the water passing in unit time has momentum aDv^2 along CD, while after impact the momentum of the water is in the direction DB. Since the reaction of the plane, if we neglect friction, must be normal to the plane, or along DH, we have, if DE represents the momentum of the jet before impact, that the momentum which has to be added to produce the actual momentum along DF is represented in magnitude and direction by DH, where DEFH is a parallelogram. Thus the force exerted by the jet on the plane is represented by DG, which is equal and opposite to DH. The force acting on the plane may be resolved into

a component DL perpendicular to direction of the jet, and a component DI parallel to the original direction of the jet. We have above considered the case of a jet of liquid; a similar effect would be produced if a current of air were striking the plane in the direction CD, when we should be dealing with the case of a kite. The upward component of the pressure on the plane supplies the lift which counteracts gravity. The function of the string is to counteract the horizontal component. Owing to the fact that the string is not horizontal but slopes downwards, the tension in the string has a vertical component, so that the upward component of the pressure of the wind on the kite has to neutralise not only the weight of the kite but also this component. In an aeroplane, in place of the wind moving and the surface being at rest, the air is at rest and the surface is in movement. The result, as far as the plane is concerned, is exactly the same. The component DE, called the drift, is neutralised by the thrust of the propeller, while the component DL, called the lift, neutralises the attraction of gravity.

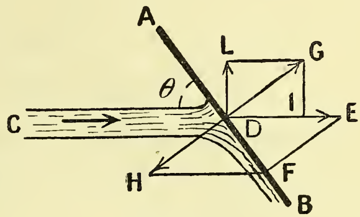


FIG. 82.

CHAPTER X

MOLECULAR PHENOMENA IN LIQUIDS

52. Surface Tension. Capillarity.—In the preceding chapters we have considered the effect of external forces on bodies, but have in general neglected the forces which one part of a body may exert on the neighbouring parts. We have now to describe certain phenomena which depend on the forces which are in play between the different molecules of which we may consider a liquid to be composed. That the molecules of a liquid do exert an attraction on each other can be exhibited by an experiment due to Osborn Reynolds. If a glass tube about 60 inches long and closed at one end is very carefully cleaned, and, the inner surface having been wetted with sulphuric acid, the tube is filled with mercury and then inserted over a vessel containing mercury, the mercury at first falls, giving a Torricellian vacuum. If, however, the tube is lowered so as to be filled by the mercury and the small bubble of air generally found is removed, the operation being repeated if necessary several times, it will be found that the tube can be raised to the vertical position and the mercury will still *completely* fill the tube. Since the atmospheric pressure will only support a column of about 30 inches in height, it follows that the mercury at the top of the tube is under a considerable tension, and yet the column does not break owing to the molecules of the liquid cohering. The tension which a liquid is able to sustain, when carefully freed from air, is very considerable, amounting in the case of sulphuric acid to about 170 lbs. per square inch. This cohesion of a liquid indicates that the molecules which constitute the liquid attract each other, so that a considerable force is necessary to separate them. It appears, however, that this attraction between two neighbouring molecules is only appreciable when these molecules are *very near together*. Thus the radius over which the attraction due to a molecule of water is appreciable is only about a millionth of a millimetre.

Liquids can support a tensile stress.

When we plunge even a thin knife-blade into water, the molecules on one side are too far off to exert any molecular attraction on the molecules on the other, and hence there is no resistance offered by molecular attraction to the separation. In Osborn Reynolds's experiment we have to tear the molecules apart, and at first when they begin

to separate the molecular attraction between molecules on opposite sides of the tear is quite appreciable.

If a globule of oil is floated in a mixture of alcohol and water of exactly the same density as the oil, the oil assumes a spherical shape. By immersing the oil in a liquid of the same density the influence of gravity is removed, and hence the molecules of the oil under the attractive force they exert on one another arrange themselves in that form in which they get nearest together, that is, they form a sphere. Now exactly the same result would occur if we suppose that in place of the molecules exerting the attraction we have postulated above, the oil were contained in an elastic membrane or skin—and it is often convenient to speak as if such an elastic skin existed at the boundary of a liquid—when the phenomena observed are said to be due to *surface tension*. The manner in which surface tension phenomena are produced by the molecular attraction can be seen if we consider the effect of the attraction of neighbouring molecules. If we describe a sphere about any given molecule of which the radius is equal to the range of molecular attraction, it is only

Surface tension of a liquid.

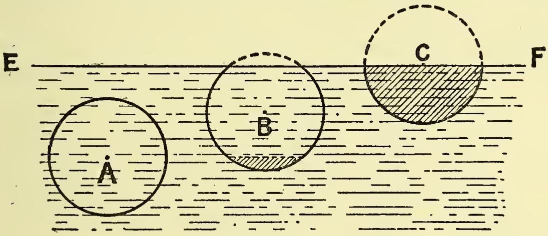


Fig. 83.

molecules which lie within this sphere which will exert force on the given molecule. In the case of a molecule A (Fig. 83) well within the liquid, the whole sphere will lie within the liquid, and hence the molecule A will be attracted by the neighbouring molecules equally in all directions. If, however, the molecule (B) is so near the surface of the liquid, EF, that the sphere would intersect the surface, then the attraction exerted on the molecule is not the same in all directions. The attraction due to molecules within that portion of the sphere in the liquid which is unshaded, being symmetrical about the molecule B, will have a resultant which is zero. The attractions of the molecules within the shaded part will, however, have a resultant directed towards the inside of the liquid mass, and perpendicular to the surface. In the case of a molecule actually on the surface, as at C, this resultant is a maximum. The effect of these unbalanced molecular forces acting on the molecules near the surface is to exert a pressure on the interior of a liquid mass, similar to that which would be caused by an elastic skin.

The magnitude of the pressure due to the surface tension depends on the form of the liquid surface, and we may obtain an idea of the direction in which the form of the surface will effect the pressure from the following considerations. Let us take the case of three molecules, A, B, and C (Fig. 84), at equal distances, less than the radius of molecular attraction, from the surface EF, which in the first case is concave, in

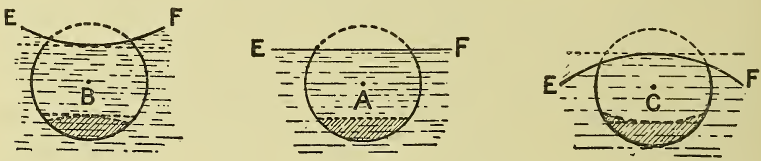


Fig. 84.

the second plane, and the third convex, and, as before, let us indicate by shading the part of the sphere of molecular attraction which is efficacious in producing an inwardly directed force on the molecule. If the surface is concave as at B, then, although the molecule B is at the same distance below the surface as is A, where the surface is plane, the shaded part is less, so that the molecular force acting on B towards the inside of the liquid is less than that on A. In the case where the liquid surface is convex (C), the shaded part is larger than in A, and hence the force is larger. Looking at it from the point of view of an elastic membrane, it is evident that at B the elasticity of the membrane would diminish the pressure within the liquid, while at C it would increase the pressure.

The existence of this pressure due to molecular, as distinct from gravitational, attractions cannot be directly demonstrated by experiment, but there are many striking phenomena depending on the fact that the surface of a liquid is in a state of tension. Thus if a metal ring is dipped in a solution of soap, and a small loop of cotton, which has been previously moistened with the solution, is placed on the film left on the ring, this loop can be made to take up any form such as A (Fig. 85), and will retain this form. If, however, the film *within* the loop is broken, the loop immediately takes up the circular form shown at B; and if it is now deformed in any way, on being released it immediately springs back to the circular form. This behaviour is due to the fact that, in the first case, the

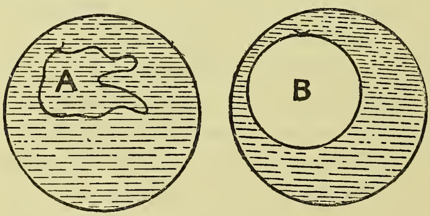


Fig. 85.

surface tension of the liquid film acts equally on both sides of the cotton, but when the film inside the loop is broken, the surface tension only acts on one side, and hence draws the loop out into a circle. Another method of showing the surface tension is by means of a bent wire ABC (Fig. 86) and a straight wire DE, which simply rests against this. If a soap film is formed in the enclosed space DBE, it will be found that the surface tension acting on DE is able to support not only the weight of the wire DE, but also a small weight W. This arrangement might also be used to obtain a rough measure of the amount of the surface tension. If W is the mass of the cross wire DE and its attached weight, then the surface tension of the film supports weight W, and therefore exerts a force of Wg units of force. The surface tension of the film acts all along the portion of the wire DE, intercepted between the legs of the bent wire, and acts at right angles to the wire. Since the film has two surfaces, if the force exerted on unit length of DE due to the surface tension of *one* side of the film be T, then the whole upward force on DE due to surface tension is $2Tl$, where l is the length of DE in contact with the film. Hence if there is equilibrium

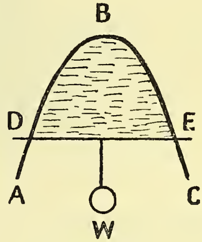


FIG. 86.

$$2Tl = Wg,$$

or
$$T = \frac{Wg}{2l}.$$

The quantity T is called *the* surface tension of the liquid, and is the force exerted across unit length taken along the surface of the liquid. In the C.G.S. system the surface tension is measured in dynes per centimetre. Owing to surface tension, work has to be done to stretch a film, and it can be shown that to increase the area of the surface of a liquid by unity the work which has to be done against the surface tension is numerically equal to T.

If a plate of glass with its sides vertical is plunged in water, it will be found, as shown at (a), Fig. 87, that where the liquid touches the glass it is drawn up above the level of the general surface. If, however, the glass is placed in mercury, the surface of the liquid near the glass is depressed below the general surface, as shown at (b). The angle BPA between the tangent to the liquid surface at the point P, where it meets the solid, and surface of the solid is called the *angle of contact* between the liquid and the solid. The angle of contact between a solid and a liquid depends on the third material, which exists above the free surface of the liquid. Thus the angle of contact between mercury and glass, when air is above the

Angle of contact.

mercury, is different from the angle of contact when there is a layer of water above the mercury. In the case where the angle of contact APB is less than 90° ((*a*), Fig. 87), the surface tension of the liquid surface supports the part of the liquid which is above the general level. In the

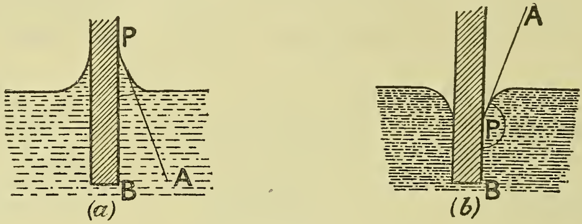


Fig. 87.

same way the surface tension, when the angle of contact is greater than 90° ((*b*), Fig 87), withstands the hydrostatic pressure due to the liquid displaced near the surface of the solid. In the case of a liquid which "wets" the surface, as happens with water and clean glass, the angle of contact is zero, and in such a case a drop of the liquid placed on a horizontal plate will spread all over the plate. If the angle of contact is greater than 90°, under similar circumstances the drop will remain all huddled together with the edge tucked in in the manner familiar in the case of a drop of mercury on a table.

Ascent of a liquid in a capillary tube.

If the end of a fine clean glass tube, called a capillary, is dipped in water, the water will rise in the tube so that the miniscus in the tube is at a higher level that the water outside. This effect is due to surface tension, and by measuring the rise the magnitude of the surface tension can be calculated. If h (Fig. 88) is the height to which the water is raised, then the weight of the liquid which is supported by the surface tension is $\pi r^2 hgD$, where r is the radius of the tube and D is the density of the liquid. The liquid meets the solid at the meniscus E along a circle of circumference $2\pi r$, and hence, if the surface tension is T, the force exerted by the liquid on the glass, or the glass on the liquid, is $2\pi rT$. Since the liquid is supposed to wet the glass so that the angle of contact is zero, the direction of the reaction of the glass on the liquid at the surface is upwards and parallel to the axis of the tube. Since it is this reaction which supports the weight of the column of liquid, we must have

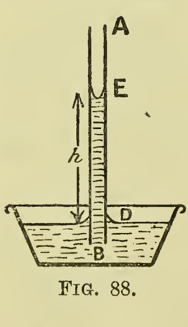


Fig. 88.

$$2\pi rT = \pi r^2 hgD$$

or

$$T = \frac{rhgD}{2} \quad . \quad . \quad . \quad . \quad (40)$$

If the angle of contact in place of being zero has a value θ, the resolved part of the reaction parallel to the axis of the tube will be $2\pi rT \cos\theta$, and hence

$$T = \frac{rhgD}{2\cos\theta} \quad . \quad . \quad . \quad . \quad (41)$$

It has been found that the surface tension of a water surface which is contaminated with oil is very considerably less than that of a surface of pure water. This fact explains the action of oil in quieting sea waves. The waves are caused by the wind acting on the surface of the water, and it appears that the "grip" of the wind on the surface depends in a great measure on the surface being ruffled or corrugated with small waves or ripples, and hence any agent which prevents the formation of these ripples will reduce the power of the wind to raise large waves. When a small wave is formed on the surface of water the surface is stretched, for obviously the wavy surface has a greater area than a plane surface. Owing to the stretching of the surface the oil film is made thinner, and hence the surface tension is increased, this increase of surface tension tending to oppose the production of the wave. Thus although the variation in surface tension is quite a small force, yet by preventing the formation of the ripples it prevents the wind obtaining a purchase, and hence is able to exercise a very marked effect on the production of the large waves met with in the sea.

Effect of oil on waves.

53. Viscosity.—If a vessel, such as a tumbler, containing water in which some light powder has been scattered is rotated about its vertical axis, it will be found that while the layer of water in immediate contact with the wall starts moving at once, the rest of the water remains at first at rest. As, however, the rotation is continued the motion is communicated to the water nearer and nearer to the centre, and finally, after a little time, the whole of the water rotates at the same speed as the glass. The reason the water at first remains at rest is of course its inertia, and the fact that after a time it becomes set in motion shows that some external force must have been in action. This force is due to the fact that if two layers of a liquid are moving at different speeds, then the faster moving layer experiences a resistance to its motion, while the slower moving layer experiences a force tending to increase its velocity. This effect, which is analogous to friction between solid bodies, is called, in the case of fluids, *viscosity*.

If we consider two small parallel plane areas a taken parallel to the direction of motion of and moving with the fluid and at a distance d apart, and if the difference in the speed with which the fluid is moving at these two areas is v, then the resistance R experienced by the faster moving area will be given by

$$R = \frac{va}{d} . c \quad . \quad . \quad . \quad . \quad (42)$$

where c is a constant which depends on the nature and temperature of the liquid, and is called the coefficient of viscosity of the liquid. The slower moving area will experience a force in the *direction of motion* also equal to R.

When a liquid flows through a tube it has been proved that the layer of liquid in actual contact with the wall of the tube remains at rest, the velocity increasing as we go from the wall to the axis of the tube. Hence owing to viscosity the motion of the liquid along the tube is resisted, and work has to be performed to drive the liquid through the tube. The volume of liquid, for which the coefficient of viscosity is c, which will flow through a *long narrow* tube when a given difference of pressure, p, exists between the two ends is given by

Resistance to flow of liquid through a tube.

$$V=\frac{\pi p r^4}{8lc} \quad . \quad . \quad . \quad . \quad (43)$$

where r is the radius and l the length of the tube. Since the volume of liquid which passes varies as the fourth power of the radius of the tube, the amount of liquid which will pass falls off very rapidly as the size of the tube is diminished. Thus in the case of the carburetter of a petrol engine, where a fine tube or jet is used to regulate the flow of fuel, a very small change in the size of the jet will make a comparatively large alteration in the flow of petrol.

It will be noticed that owing to viscosity work has to be done when a liquid is sheared (§ 36), for the viscous resistance acts as a force which opposes the shear. This effect, however, is only produced while the shear is *altering*, so that it is not opposed to the statement previously made that a liquid can offer no permanent resistance to a shear.

The frictional resistance between two well-lubricated surfaces is chiefly due to the viscosity of the layer of lubricant. This fact is often very noticeable in the case of a motor-car engine, for when the engine is cold it requires a considerable effort to turn it. When, however, the engine is hot the viscosity of the oil is very much reduced, so that it becomes comparatively easy to turn the engine.

54. Diffusion of Liquids. Osmosis.—If two liquids which are miscible are introduced into a vessel so that the denser is below and the lighter above, it is found that some of the lighter liquid will slowly travel down and mix with the heavier, while some of the heavier will travel up and mix with the lighter. This process of mixing which occurs in opposition to gravity is called *diffusion.*

The rapidity with which different liquids diffuse is very different. Thus substances such as hydrochloric acid and salts of the mineral acids diffuse comparatively rapidly, and are called *crystalloids.* On the other

hand, substances such as gum, starch, albumen, caramel, and the like, which are glue-like amorphous bodies, diffuse very slowly, and are called *colloids.*

If a film of a colloid, such as paper coated with starch, or a sheet of parchment, is used as a partition in a vessel, and pure water is placed on one side and a solution of crystalloids and colloids is placed on the other, it is found that the crystalloids will diffuse through into the water but the colloids will not. Thus a colloid membrane is permeable to crystalloids but impermeable to colloids, and hence such a membrane is called a semi-permeable membrane, for it is permeable to one class of bodies but not to the other class.

Another characteristic of a semi-permeable membrane is illustrated by the following experiment.

If in a vessel A (Fig. 89), such as a thistle funnel with its larger end closed by a sheet of parchment, we place a solution of copper sulphate, filling the vessel up to about D, and then place it as shown in the figure, so that the parchment is below the surface of some pure water contained in vessel C. Then it is found that the water makes its way through the parchment partition into A, the solution inside gradually rising up in the tube DB. Thus the water has been able to pass through the parchment in opposition to the hydrostatic pressure due to the column of liquid BD. After a time the water ceases to force its way through the partition, its tendency to do so being counterbalanced by the hydrostatic pressure. It will also be noticed that in time some of the copper sulphate travels out into the surrounding water. If, instead of placing the vessel A containing the copper sulphate solution in pure water, it is placed in a solution of copper sulphate of the same strength as that inside, no change in the quantity of liquid in the vessel takes place. If, however, it is placed in a stronger solution, water will pass out from the vessel A, so that the solution inside becomes more concentrated. These phenomena are called *osmosis*, and the final pressure produced in the vessel containing the salt solution, when placed in pure water, is called the *osmotic pressure.*

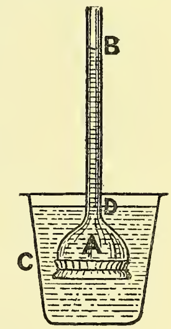


FIG. 89.

Osmotic pressure.

The osmotic pressure developed with some semi-permeable membranes is quite large. Thus a 6 per cent. solution of sugar may give an osmotic pressure of about 4 atmospheres.

We may account for the production of the osmotic pressure if we suppose that the semi-permeable membrane is porous, and that the pores are of such a size that while they allow the water molecules to pass they are too small to allow of the larger sugar molecules to pass. Thus if the

molecules of the liquid are in continual movement when a water molecule strikes the membrane it can pass through, but a sugar molecule striking the membrane is not allowed to pass. Since on the sugar side of the membrane there are fewer molecules of *water* per unit volume, fewer water molecules will strike the membrane, and hence pass through, on the solution side, than strike (and pass through) on the pure water side. Hence an accumulation of water takes place on the solution side, and this accumulation produces the osmotic pressure.

In the case of both animals and plants there exist cells of which the walls are semi-permeable membranes, and hence the phenomena of osmosis plays a prominent part in the physiological processes which go on in living tissue.

CHAPTER XI

PROPERTIES OF SOLIDS

55. Homogeneous and Isotropic Bodies.—By a *homogeneous* body is meant one which has the same physical properties throughout. A body such that when a spherical portion is tested in different *directions* it exhibits no differences in any physical property is said to be *isotropic.* Except under very exceptional circumstances all liquids and gases are isotropic. Some solids, however, notably crystals, while being homogeneous, have different physical properties in different *directions.* Unless otherwise stated, we shall confine our attention to isotropic solids.

We may define a solid as a body which is capable of offering *permanent* resistance to a shear. As will be seen later, this definition in the case of all actual solid bodies has to be limited, for if the shear exceeds a certain amount, which varies for different substances, the resistance offered, that is the stress, decreases and finally the body will rupture, when although the shape will have altered yet there will be no stress. Hence, unless otherwise mentioned, we shall suppose that the strain is in every case so small that on the removal of the deforming force the body regains its original shape.

The behaviour of some solids differs very greatly according to the rate at which the strain is produced. Thus if the finger is pressed against a cake of cobbler's-wax, the substance will behave like a very viscous liquid, so that the cake can be indented with the finger. If, however, the cake is given a *sudden* flick with the finger it will be shattered, breaking up just like a piece of glass. Hence for a slowly applied deforming force the wax is practically a liquid, but to a quickly applied force it behaves as a very brittle solid.

56. Determination of the Density of a Solid.—The density of a solid is generally obtained by measuring the weight W and also the volume V, when the density is given by W/V. In all cases the determination of the weight is simple, a balance being employed. Different methods are employed for measuring the volume.

The most usual way of measuring the volume is to apply the principle of Archimedes. Thus if we determine the decrease in the apparent weight of a body when immersed in a liquid of density D, we know that this loss of weight is equal to VD, where V is the volume of the body. Hence if we know D we can immediately calculate V, and thence deduce the density of the body. In the case of a body such as sand,

which is in such small particles that it is impracticable to suspend it so as to apply the above method, we make use of the fact that when immersed in a liquid it displaces its own volume of the liquid, using the specific gravity bottle described in § 47. By weighing the bottle first empty and then full of water we determine its volume. It is then emptied and partly filled with sand, the difference between the weight now and when empty giving the weight of the sand. Then, leaving the sand in, the bottle is filled with water and again weighed. The difference between this weight and the weight when filled with water only gives the weight of the sand, less the weight of the water displaced. Hence the weight of water displaced is obtained, and knowing the density of the water we can calculate its volume, which must be equal to the volume of the sand.

57. Elasticity of Solids. Young's Modulus.—As is the case with liquids, the bulk elasticity of solids is very large, and has to be taken into account when considering the strains set up in bodies of many shapes when they suffer deformation under the action of applied forces, for such strains often consist of the combination of a volume strain and a shear. The study of such combined strains is in general beyond the scope of this work; we may, however, consider one particular case, namely, that of a stretched wire.

If to a wire of length L and radius r we apply a longitudinal pull P, it was first shown by Hooke that the elongation, l, produced is proportional to P as long as P is not so great as to cause a permanent elongation.

When a wire is stretched both the shape and volume are altered, so that both types of strain are produced; since, however, the particular combination found in the case of the wire is of considerable practical importance, the ratio of the stress to the strain, as measured by the elongation, has received a special name, and is called Young's modulus. The stress in this case is the applied tension per unit cross-section, or $P/\pi r^2$, while the strain is the elongation per unit length of the wire, or l/L. Hence Young's modulus Y is given by

Young's modulus.

$$Y = \frac{\text{stress}}{\text{strain}} = P/\pi r^2 \div l/L = \frac{PL}{\pi r^2 l} \qquad . \quad . \quad (44)$$

When a rod AB (Fig. 90) is held firmly fixed at one end A and a force is applied at the other end at right angles to the length, the rod becomes bent into the form AB′. The material situated along the central line, or neutral axis as it is called, which is shown dotted, is neither extended or compressed owing to the bending. All parts of the rod above this axis have undergone a stretching, the amount of stretch increasing as we go away from the axis. All parts below the

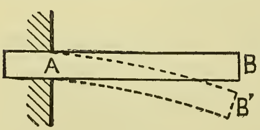


Fig. 90.

axis have undergone compression, the amount of the compression being greater the greater the distance from the axis. Thus although the strain in different parts of the rod varies, it is everywhere of the same type as that which occurs when a wire is stretched, and it can be shown that the amount of bending produced by a given force is proportional to Young's modulus for the material. If the length of the rod projecting beyond the support is L, and the cross-section is a rectangle of breadth b and depth d, and l is the distance through which a force P will deflect the end, then

$$Y = \frac{4PL^3}{bd^3 l} \quad . \quad . \quad . \quad . \quad (45)$$

This expression shows that the stiffness of such a rod varies directly as the breadth, directly as the *cube* of the depth, and inversely as the *cube* of the length.

Torsion of a wire.

If one end of a wire of radius r and length L is kept fixed while a twisting couple u is applied to the other end, then so long as the couple is not too great, the angle θ through which the end of the wire will be twisted is proportional to the applied couple. It can be shown that in this case, though the amount of strain produced in the material varies with the distance from the axis of the wire, yet everywhere it is a pure shear, and that if n is the rigidity modulus, then

$$\theta = \frac{2Lu}{\pi r^4 n} \quad . \quad . \quad . \quad . \quad (46)$$

It will be noticed that the twist produced by a given couple varies inversely as the *fourth* power of the *radius*. Thus if we halve the radius of the wire, the twist produced by a given couple will be increased sixteen (2^4) fold. As we shall see later, fine wires are often used to suspend bodies, such as galvanometer needles, in which we require a measurable deflection to be produced by a very small couple. Hence the importance of using very fine wires or fibres in such cases.

58. Elastic Limit. Yield Point. Breaking Stress.—When considering the elasticity of solids, we have supposed that the strain to which the specimen has been subjected has been so small that on the removal of the deforming force the body has returned to its original condition. If, however, the strain is gradually increased, a point will at length be reached when on the removal of the force the body does not completely recover its original shape or volume. The strain, or what is more often specified the corresponding stress, below which the recovery is complete, is called the elastic limit for the body.

Elastic limit.

In the case of a rod of a material such as steel to which a gradually

increasing tension is applied, the relation between the tension and the elongation produced is as indicated by the curve OABCD (Fig. 91). The part of the curve from O to A is straight, indicating that the elongation is proportional to the tension, *i.e.* that Hooke's law holds, A being the elastic limit. Between A and B the elongation increases faster than the tension, while at B a more or less sudden increase in the elongation is produced without the tension being much increased. The point B is called the *yield point* of the material. After the yield point the curve rises slightly and reaches a maximum at C, at which point the elongation goes on increasing although the tension is kept constant. Beyond C the curve falls rapidly, indicating that even

Yield point.

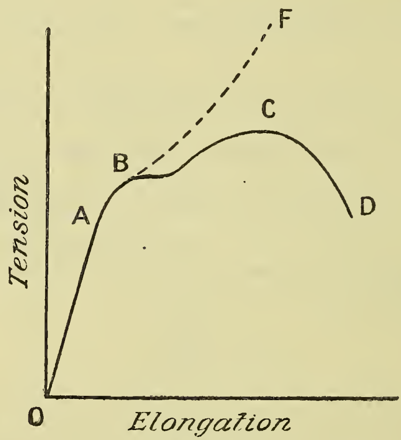


FIG. 91.

with a decreasing tension the elongation increases, till finally at D the specimen ruptures.

Up to the point B the cross-section of the specimen only decreases slightly, while up to C the decrease in section is uniform throughout the length. At C, however, a local contraction begins to appear at one place in the specimen, and this contraction rapidly increases till the specimen breaks. The tensions considered above are not true stresses, since the cross-section varies. If allowance is made for this variation, the stress-strain diagram would follow the dotted line BF. We cannot continue it beyond F, since after this the elongation is not uniform throughout the length of the specimen, and hence we cannot calculate the elongation per unit length.

If when the specimen has reached any condition beyond that indicated by A on the diagram the tension is reduced to zero, the specimen

does not return to its original length, indicating that a permanent deformation has been produced.

Breaking stress.

The quotient of the maximum tension, *i.e.* that corresponding to the point C, by the original cross-section of the specimen is called the breaking stress of the specimen. The percentage increase in length in a given length, including the point of fracture and the percentage decrease in area of the cross-section at the fracture, are generally recorded when testing specimens of materials for engineering purposes.

BOOK II

HEAT

CHAPTER I

TEMPERATURE AND EXPANSION

59. Temperature. Thermometers.—The qualitative meaning of the term temperature is familiar to every one; thus a body which feels hot to the touch is said to have a higher temperature than a body which feels cold. Our senses, however, do not allow of our making any but the very roughest estimate of the *amount* by which the temperature of one body is higher than that of the other. Hence in order to measure temperature we are obliged to make use of the change in some physical property of some kind of matter which takes place as the temperature changes. The physical property which is most usually employed for this purpose is the volume of a liquid or of a gas, both of which depend on the temperature. In order to define certain fixed temperatures we also have recourse to the fact that a certain kind of matter, namely water, may according to its temperature exist as a solid, a liquid, or a gas, and that the change from one state to the other, other conditions remaining constant, always takes place at the same temperature.

Thermometric fixed temperatures.

If a mass of pounded ice is heated it will, when the temperature reaches a certain value, commence to melt, and if the mixture is well stirred, the temperature will remain constant till the whole of the ice is melted, and this temperature is taken as one of the fixed points used in measuring temperature. If the water formed by the melting of the ice is further heated it will in time commence to change into a gas, or boil as it is called, and again the temperature will remain constant till all the liquid water has been converted into the gas steam. This second temperature is taken as a second fixed temperature.

In order to subdivide the interval of temperature between these two points, use is made of the change in pressure[1] required to keep the volume of a given quantity of the gas hydrogen constant when its temperature is changed. Thus a temperature such that the increase in pressure of a given mass of hydrogen confined in an inextensible envelope,

[1] Sometimes the change in *volume* of a given quantity of hydrogen when the pressure is kept constant is taken as the measure of a change in temperature. The two scales are practically identical, the pressure scale being chosen for experimental reasons, namely, that it is in general easier to measure a change in pressure than a change in volume.

when heated from the temperature of melting ice to this temperature, is half the increase in pressure when the heating is continued to the temperature of boiling water, we *call* a temperature half way between that of melting ice and boiling water. When making the above convention we tacitly assume that the rate of change of pressure of hydrogen at constant volume with temperature is constant. We shall see, however, that the *rates* of increase of pressure with temperature for different gases are decidedly different, and hence the above assumption as to hydrogen being the *one* gas for which the rate of increase in pressure is proportional to the temperature requires justification. This point, however, must for the present be postponed, but we shall see that the scale of temperature based on the expansion of hydrogen does agree very nearly with another scale which is independent of the properties of any particular substance.

Change of pressure of hydrogen taken as a measure of temperature change.

There are several ways of subdividing the temperature interval between the fixed points, called scales of temperature, in general use. The most usual one is to call the temperature of melting ice zero or 0°, and that of boiling water 100°, the interval being divided into a hundred parts called degrees. This scale is called the *centigrade* or Celsius scale, and is almost exclusively used in scientific measurements. Another scale, called the *Fahrenheit* scale, takes the temperature of melting ice as 32° and that of boiling water as 212°, the interval being divided into 180 degrees. Finally, in the *Réaumur* scale the temperature of melting ice is called 0°, and there are 80 degrees between this and the temperature of boiling water. Since 0° C. = 32° F. = 0° R., and an interval of 100° C. is equal to an interval of 180° F. and of 80° R., if c, f, and r are the readings corresponding to a given temperature on the three scales, we have

Different scales of temperature.

$$c : f-32 : r :: 100 : 180 : 80$$

or

$$\frac{c}{5}=\frac{f-32}{9}=\frac{r}{4} \quad . \quad . \quad . \quad . \quad (47)$$

and by means of this equation any temperature of one scale can be converted into either of the other scales.

In order to measure temperatures above and below those corresponding to the fixed points, the scales are prolonged upwards and downwards, temperatures below the zero being indicated by a negative sign.

Instruments to measure temperature are called thermometers. Since the use of a thermometer in which the expansion of hydrogen is employed is very inconvenient, and it requires very great skill in its manipulation, it is usual to employ thermometers in which the expansion of a

liquid is employed to indicate changes of temperature, and the corrections which have to be applied to reduce temperatures measured with this liquid to the hydrogen scale are determined. The liquid almost exclusively used is mercury, and it is contained in a glass envelope. Owing to the fact that the glass expands with rise of temperature, the apparent increase in volume of the mercury is not as great as the true increase. It is the apparent increase which, however, is used, and hence, since all kinds of glass do not expand at the same rate, the corrections required to reduce the readings of a mercury thermometer to those of the hydrogen thermometer vary slightly from one thermometer to another. For most practical purposes, however, this difference between different mercury thermometers is too small to cause any appreciable error, and even the difference between a mercury and a hydrogen thermometer can in many cases be neglected. Thus the maximum difference between a mercury thermometer of Jena glass (16^{111}) and the hydrogen thermometer between 0° and 100° C. amounts to 0·12° C.

The mercury thermometer.

The description of a hydrogen thermometer is for the present postponed (see § 64). A mercury thermometer consists of a bulb, which is generally cylindrical, containing the mercury, and connected to a fine capillary tube. The far end of this tube is closed and the space above the mercury is in general a vacuum, though for mercury thermometers to be used at comparatively high temperatures this space is filled with nitrogen at a pressure of two or three atmospheres. The stem of the thermometer is graduated so that when the bulb and the stem is raised to some given temperature, the mercury meniscus is opposite the graduation corresponding to this temperature.

Since mercury freezes at a temperature of about − 40° C., thermometers for use at low temperatures are filled with alcohol, a liquid that does not freeze till a temperature of about − 110° C. is reached. The disadvantage of alcohol is that it boils at about 80° C., and hence an alcohol thermometer cannot be used at a temperature much above 70°. For this reason toluene, which can be used over the range − 80° C. to 100°, is to be preferred to alcohol.

Errors of mercury thermometers.

Although for many purposes a mercury thermometer is a very convenient instrument for the measurement of temperature, yet where a very accurate value of the temperature is required, so many corrections have to be applied and such precautions have to be taken that its use becomes a matter of considerable difficulty. We have already referred to the fact that the scale of a mercury thermometer depends to a certain extent on the kind of glass of which it is constructed. Another effect due to the glass is a gradual rise in the readings corresponding to any given temperature, which goes on

for years after the thermometer is made. This rise is due to the gradual recovery of the glass from the effect of the heating to which it was subjected during the manufacture of the thermometer, and is called the secular rise of the zero, since it is usually detected by noting the change in the zero reading when the thermometer is placed in melting ice. In addition, a temporary lowering of the zero takes place after the thermometer has been raised to even a comparatively low temperature such as 100° C., due to the fact that the glass takes some time to recover from the expansion produced by the rise in temperature.

Determination of the fixed points of a thermometer.

The lower fixed point of a thermometer is determined by surrounding the bulb with pounded ice mixed with *distilled* water. The reason why distilled water must be used is that if there are salts dissolved in the water the temperature will be lower than if pure water is employed (§ 73).

To determine the upper fixed point, the thermometer T (Fig. 92) is suspended in a double-walled jacket A, called a hypsometer. The steam from boiling water in the can C is passed up the tube B, past the thermometer, and then down the outer part of the jacket, finally escaping through the tube D.

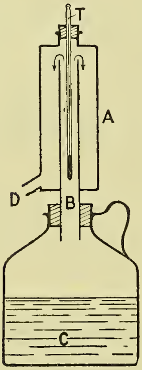


FIG. 92.

It is only when the atmospheric pressure is a standard atmosphere (§ 41) that the boiling-point of water is 100° C., hence when determining the upper fixed point of a thermometer the barometric height must be observed, and a correction must be applied if the pressure has not the standard value. This correction may be obtained from a table giving the boiling-point of water at different pressures.

The reason for placing the thermometer in the steam and not in the boiling water is that it is found that the temperature of the boiling *water* depends to a certain extent on the nature of the vessel in which the water is contained. Another important consideration is that small quantities of impurity in the water alter the temperature of the water at which ebullition takes place, but do not affect the temperature of the *steam*.

Maximum thermometers.

It is sometimes necessary to know the highest temperature to which the thermometer has been exposed, and for this purpose what is called a maximum thermometer is employed. The ordinary clinical thermometer is such a maximum thermometer in which there is a slight constriction in the bore near the bulb. When the temperature rises the mercury is forced past the constriction, but when the temperature falls the thread breaks at the

constriction, and hence the mercury left in the stem indicates the highest temperature reached. To reset the thermometer the mercury in the stem is jerked down past the constriction, or centrifugal force is used to perform this action, the thermometer being rapidly swung in a circle with the bulb outwards. Since, as was shown in § 19, the acceleration acting on a body which moves in a circle of radius r is ro^2, where o is the angular velocity, so that the centrifugal force is mro^2, it is obvious that it is of advantage to swing the thermometer through an arc of as large a circle as possible, *i.e.* to swing it at arm's length, and also to make the angular velocity of the rotation as great as possible.

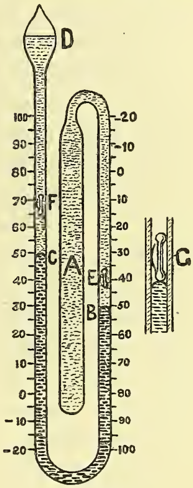


Fig. 93.

The form of maximum thermometer generally used in meteorology is Rutherford's, and consists of an ordinary mercury thermometer with a small iron index fitting loosely in the bore. The thermometer is generally placed in a nearly horizontal position, and when the temperature rises the mercury column drives the iron index forward, since mercury does not "wet" iron. When the temperature falls the index remains, and the end *next* the bulb indicates the highest temperature reached. Rutherford's minimum thermometer consists of an alcohol thermometer with a glass index placed inside the liquid column. When the temperature falls the surface tension draws the index along, the far end of the index touching the meniscus. When the temperature rises the alcohol flows past the index, so that the end of the index *furthest* from the bulb indicates the lowest temperature reached. In Six's thermometer the maximum and minimum temperatures are shown by two small indexes in the same thermometer. The thermometric liquid is alcohol contained in the bulb A (Fig. 93). At the end of the alcohol column is a thread of mercury BC, while the remainder of the tube and *part* of the bulb D is filled with alcohol. The two ends of the mercury thread serve to indicate the temperature on scales placed alongside. Two small glass rods, each fitted with a steel spring, as shown at G, are placed in the alcohol, and indicate the highest positions reached by the ends C and B of the mercury. Thus the bottom of the index F shows the maximum temperature and the bottom of the index E shows the minimum temperature to which the thermometer has been exposed.

Minimum thermometer.

The description of other instruments for the measurement of temperature is postponed till we come to the discussion of the particular physical property on the variation of which with temperature they depend for their action.

60. Expansion of Solids. Coefficient of Expansion.—When the temperature of a solid body is raised, in general the distance between any two points in the body increases, that is, the body expands. Thus a cylindrical rod of iron when the temperature rises increases not only in length but also in diameter. If we only consider the increase in *length*, we are said to deal with the *linear expansion* of the body. If, however, the change in *volume* is considered, we are said to deal with the *cubical expansion* of the body.

The expansion of solid bodies with heat is made use of in many ways. Thus the iron tyres fitted round wooden cart-wheels are made slightly smaller in diameter than the wooden rim. The tyre is then heated, so that it expands, and thus can be placed round the wheel. When cool the iron contracts again and binds the wood firmly together. In all large metal structures, such as bridges, very careful provision has to be made to allow for the changes in length of the different parts with change in temperature, otherwise strains will be set up which might produce rupture.

Coefficient of linear expansion.

If a bar of a solid is heated from 0° C. to 1° C., the ratio of the increase of length to the original length is called the *coefficient of linear expansion*. Thus if the length of a bar at 0° C. is L_0 and the coefficient of linear expansion is a, the length at a temperature of 1° C. will be $L_0 + L_0 a$, and the length L at a temperature t will be given by

$$L = L_0(1 + at) \quad . \quad . \quad . \quad . \quad (48)$$

The length at a temperature t' will be $L_0(1 + at')$, and hence the increase in length when heated from t to t' will be $L_0(1 + at') - L_0(1 + at)$ or $L_0 a(t - t')$.

Suppose we are given the length L at a temperature t, and wish to calculate the length at a temperature t'. We have that $L_0 = L/(1 + at)$, and hence

$$L' = \frac{L(1 + at')}{1 + at} \quad . \quad . \quad . \quad . \quad (49)$$

Since $\frac{1}{1 + at} = 1 - at$, together with terms involving higher powers of a, if a is so small that we can neglect all the terms involving a^2 and higher powers,[1]

$$\begin{aligned} L' &= L(1 + at')(1 - at) \\ &= L\{1 + a(t' - t)\} \quad . \quad . \quad . \quad . \quad (50) \end{aligned}$$

where as before the term a^2tt', which involves a^2, is neglected.

[1] The smallness of such terms can at once be appreciated, for a has for steel a value ·000011, and hence $a^2 = $ ·000000000121.

This result shows that when a *is small* we can make use of an expression, equation (50), of the same form as equation (48) to calculate the length L at any temperature t' being given the length L at any other temperature t. If, however, as is the case of the cubical expansion of gases, the coefficient of expansion is not very small, equation (50) cannot be used without producing appreciable error. In such a case we must employ the expression given in equation (49). Some stress has been laid on this point, as students occasionally attempt to use an expression of the form of equation (50), where, owing to the fact that a is not very small, it cannot be applied.

The above investigation shows that in the case of the linear expansion of a solid, where a is always very small, we need not define the coefficient in terms of the length at 0° C., but may say that the coefficient of linear expansion of a solid is the increase of length of unit length when the temperature is raised 1° C.

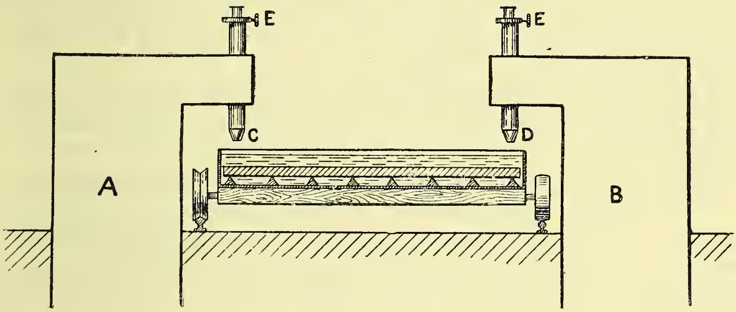


FIG. 94.

Of course, if the temperature scale used is the Fahrenheit, since a Fahrenheit degree is only 5/9ths of a centigrade degree, the coefficient of expansion will only be 5/9ths of the value when the centigrade scale is used.

Measurement of the coefficient of linear expansion.

Owing to the smallness of the increase of length to be observed, the measurement of the coefficient of linear expansion of a solid requires very refined apparatus. The principle of the method employed when a length of the material of about a metre is available is shown in Fig. 94. The bar to be tested is placed in a trough which can be filled with water, and two very fine lines which are engraved near the ends of the bar are focussed by two microscopes C and D, which are carried on massive stone pillars. These microscopes are provided with cross-wires in the eye-piece, which cross-wires are adjustable by very fine micrometer screws E. On the same carriage which supports the trough in which the bar is

placed is a second trough, which contains a standard metre scale packed round with ice. The bar having been packed round with melting ice, so that its temperature is 0°, the cross-wires are adjusted so that they coincide with the lines on the bar. The bar is then wheeled to one side and the standard metre brought under the microscopes, and the readings on the micrometer screws when the cross-wires are brought to coincide with the images of two of the divisions are noted. Then knowing the pitch of the micrometer screws, the length of the bar at 0° C. can be deduced from the change in the readings of the micrometer screws and the length of the scale between the two divisions employed. The temperature of the water in the trough containing the bar is then raised, and while the water is kept well stirred, the temperature is read on several thermometers which lie alongside the bar, and the cross-wires are again adjusted to coincide with the lines on the bar. The scale, of which the temperature has been kept at 0°, is then again placed below the microscopes, and the readings repeated, giving the length of the bar at the new temperature. From the readings thus obtained, the coefficient of expansion can then be calculated.

When only a comparatively short length of the solid are available, other methods which depend on the interference of light have to be employed.

Relation between coefficients of linear and cubical expansion.

Suppose we take a cube, of a material of which the coefficient of linear expansion is a, of which each edge is of unit length at 0° C., and heat this cube to a temperature of 1° C. As a result the length of each edge of the cube will become $1+a$, and hence the *volume* will be $(1+a)^3$. Now the increase of volume per unit volume per degree is called the coefficient of cubical expansion b. Hence we have

$$b=(1+a)^3-1=3a+3a^2+a^3.$$

But since a is very small, the terms in a^2 and a^3 can be neglected, so that

$$b=3a.$$

The coefficient of cubical expansion is therefore three times the coefficient of linear expansion.

If a body is so constrained that it cannot expand when the temperature rises, it will exert a very great force on the system which prevents the expansion. We can easily get an idea of the forces thus called into play in the case of a short stout rod.

Suppose a rod of length L and cross-section s composed of a material of which the coefficient of linear expansion is a, and for which Young's modulus is y. Then if the rod is heated through a temperature range of t, it will expand by an amount Lat. If a force P is used to compress

the rod longitudinally, we have (§ 57) $P=\frac{sly}{L}$, where l is the amount by which the rod is shortened owing to the compression. Hence if at the higher temperature the rod is to be brought back to its original length, we must put $l=Lat$ in this expression, so that

$$P=\frac{sLaty}{L}=saty.$$

For steel, $a=\cdot000011$ and $y=2\times10^{12}$ dynes per square centimetre. Hence if the cross-section of the rod is unity ($s=1$) and it is heated through 100° C. ($t=100$) we must have

$$\begin{aligned}P&=\cdot000011\times100\times2\times10^{12}\\&=2\cdot2\times10^{9}\text{ dynes}=2243\text{ kilograms}=22\cdot07\text{ tons.}\end{aligned}$$

Thus if we attempt to prevent such a rod from expanding, it will exert a force equal to about 22 tons weight, and it is obvious that if we have a structure the parts of which are held rigidly together, if the temperature of part of the structure changes, unless provision is made to allow for the expansions of the parts such large stresses may be produced as to cause fracture. In the same way if a solid body is heated, so that the temperature of some parts increases while that of other parts does not, fracture may easily result. This effect is particularly noticeable when we are dealing with a brittle material such as glass, in which the conductivity (§ 77) is low, so that neighbouring parts of the glass may reach very different temperatures. The reason a platinum wire can be fused into glass and iron cannot is that the coefficient of expansion of platinum is nearly the same as that of glass, while iron expands considerably more than glass, so that on cooling a piece of iron wire, which has been fused into a piece of glass, contracts more than the glass, and the joint is torn apart. In the case of a material such as fused silica (quartz) of which the coefficient of expansion is excessively small, a red-hot stick can be plunged into cold water without fracture, although the material is brittle.

61. Expansion of Liquids. Apparent Expansion.—In the case of liquids the only kind of expansion with which we have to do is volume expansion, or cubical expansion. When we consider the increase in actual volume of a liquid when heated, we are said to deal with the true expansion of the liquid. If, however, what we measure is how much the volume appears to increase when both the liquid and the containing vessel are heated, we are said to deal with the *apparent* expansion of the liquid. It is evident that the apparent expansion must be less than the true expansion, as when the temperature rises the volume of the containing vessel increases.

The increase in the volume of a vessel of simple form can easily be calculated. Thus consider a solid cylinder, ABCD (Fig. 95), of the material of length l when at a temperature of 0°, and let us consider what happens to a portion, EFGH, of which the radius 0° C. is r, when the temperature rises to t. The radius of this portion of the solid will become $r(1+ct)$ where c is the coefficient of linear expansion of the solid, and at the same time this length of the cylinder will become $l(1+ct)$. Hence the volume of the portion EFGH of the solid cylinder will change from $\pi r^2 l$ to $\pi r^2(1+ct)^2 l(1+ct)$ or $\pi r^2 l(1+ct)^3$, which, if c is small, is equal to $\pi r^2 l(1+3ct)$. But $\pi r^2 l$ is the volume of the cylindrical part EFGH at 0°. Hence the volume at t is the volume at 0° multiplied by $(1+bt)$, where b is the coefficient of cubical expansion of the solid, and is equal to $3c$. Now if the cylinder EFGH were removed, the outside shell would expand just as it does when

Increase in volume of a hollow vessel.

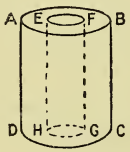


Fig. 95.

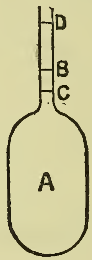


Fig. 96.

the whole cylinder is solid, and therefore the volume of the *inside* of the hollow cylinder at t is equal to the volume at 0° multiplied by $(1+bt)$. The same result can be easily obtained in the case of a spherical or a cubical vessel, and the result can be extended to a hollow vessel of any shape.

Suppose that a glass vessel A (Fig. 96) is filled with a liquid, say mercury, and that, when the whole is at a temperature of 0°, the volume of the mercury, *i.e.* of the glass vessel, is V_1; the surface of the mercury being at B. Suppose, now, it were possible to raise the temperature of the glass vessel to 1°, keeping the temperature of the contained mercury still at 0°. The vessel would expand, and its volume up to the mark B would become $V_1(1+b)$, where b is the coefficient of *cubical* expansion of the glass. Hence the mercury surface would fall to C, the volume of the bore between B and C being $V_1 b$. If, now, the mercury was also heated to 1° it would expand, and its volume would become $V_1(1+a)$, where a is the true coefficient of expansion (cubical) of the mercury. Hence if the surface of the mercury now stood at D, the volume of the

bore between C and D would be V_1a. Therefore the volume of the bore between B and D is $V_1(a-b)$. If we had not allowed for the expansion of the envelope, we should have taken the volume $V_1(a-b)$, *i.e.* the volume of the bore between the first and last positions of the surface of the mercury, as the expansion, and this volume divided by the volume at 0°, *i.e.* $V_1(a-b) \div V_1$, or $a-b$, as the coefficient of expansion of the liquid. Hence the *apparent* coefficient of expansion or dilatation is equal to the difference between the true or real coefficient of expansion (a) of the liquid and the coefficient of cubical expansion (b) of the solid envelope.

Apparent coefficient of expansion of a liquid.

The bulb shown in Fig. 94 is called a volume dilatometer, and is employed for determining the apparent expansion of liquids. The cubical expansion of the glass is obtained either by filling the bulb with a liquid of known absolute expansion, and making a series of measures, or by experiments on the linear expansion of a rod of the same glass. This known, the absolute expansion of the liquids can be calculated from the apparent expansion. The stem of the dilatometer is graduated, and the volume, up to the zero graduation, is determined by first weighing the dilatometer empty, and then when filled to the zero mark at 0° with mercury or water, and from the weight of mercury or water calculating the volume, using the known density of these liquids at 0°. The volume of the bore between two divisions is obtained in the same way by weighing a thread of mercury which, when in the stem, occupied a known number of divisions. The dilatometer is filled with the liquid, the expansion of which has to be measured, and the position of the surface in the stem noted when the dilatometer is placed in melting ice. The whole is then heated to a temperature $t°$ in a water bath, the temperature being measured by a thermometer placed in the water, and the position of the liquid surface again noted.

If V_0 is the volume at 0°, and v the increase in volume as read off on the scale, the apparent coefficient of expansion is v/tV_0, and the true coefficient of expansion is given by $a = v/tV_0 + b$.

Another form of dilatometer is the weight dilatometer. This resembles the volume dilatometer, except that there is only a single graduation on the stem. The dilatometer is first weighed empty and is then filled with the liquid and placed in melting ice. When the contents have reached a temperature of 0°, the quantity of liquid is adjusted by means of a capillary tube till the surface coincides with the mark on the stem, and the dilatometer is again weighed. The dilatometer is now heated to a temperature of $t°$, and some of the liquid withdrawn till the surface coincides with the mark, when the dilatometer is again weighed. Let w_1 be the weight of liquid *contained* in the dilatometer at the

temperature 0°, and w_2 the weight of the *contents* at a temperature t. If the dilatometer were again cooled to 0° it would contain w_2 grams of liquid, and the space between the top of the liquid and the fixed mark could hold $w_1 - w_2$ grams of liquid. Now if D is the density of the liquid at 0°, w_2 grams at that temperature would have a volume of w_2/D, and $w_1 - w_2$ grams would occupy a volume of $(w_1 - w_2)/D$. Hence a volume w_2/D of the liquid at 0° has at t appeared to *increase* in volume by an amount $(w_1 - w_2)/D$, and hence the apparent coefficient of expansion is given by

$$a - b = \frac{w_1 - w_2}{w_2 t} \quad . \quad . \quad . \quad . \quad (51)$$

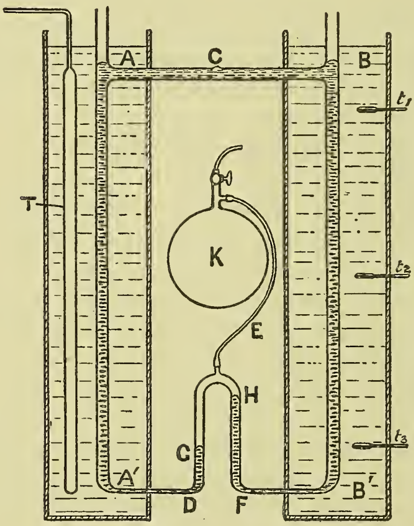


Fig. 97.

A direct method of determining the absolute coefficient of expansion of mercury was devised by Dulong and Petit. The principle of this method is to use a U-tube filled with mercury, one limb being kept at 0°, and the other at 100°. If D_0 is the density of mercury at 0°, and D_1 that at 100°, then if h_0 and h_1 are the heights of the mercury in the two limbs above the horizontal part of the U, we get by the principle of balancing columns (§ 44) $h_0 D_0 = h_1 D_1$.

Direct determination of the real coefficient of expansion of mercury.

Now the volume of unit mass of mercury at 0° is $1/D_0$, and at 100° is $1/D_1$, so that the increase in volume is $\frac{D_0 - D_1}{D_0 D_1}$. The increase divided

by the volume at 0° ($1/D_0$), and by the temperature range, gives the coefficient of expansion a. Hence

$$a = \tfrac{1}{100}\frac{D_0 - D_1}{D_1} = \tfrac{1}{100}\left(\frac{D_0}{D_1} - 1\right)$$

$$\therefore \qquad a = \tfrac{1}{100}\left(\frac{h_1}{h_0} - 1\right) \text{ or } \tfrac{1}{100}\left(\frac{h_1 - h_0}{h_0}\right) \quad . \quad . \quad (52)$$

The form of apparatus as used by Dulong and Petit was subject to numerous errors, but Regnault introduced some modifications, so that most of these were avoided, and his form of the apparatus will be described.

Two upright iron tubes AA′, BB′ (Fig. 97), are connected together near the top by a horizontal tube C. At the bottom two horizontal branches A′D and B′F are connected to two vertical glass tubes DG and FH, these tubes being joined together at the top by a T-piece, the third arm of the T being connected to a glass globe K containing compressed air. The tubes AA′ and BB′ are surrounded by cylindrical vessels, which are filled with water or ice. The water in these cylinders is kept well stirred, and the temperature in the one which is heated is measured by an air-thermometer (§ 64) T, and that of the other by a series of mercurial thermometers t_1, t_2, t_3.

The upper horizontal tube AB is pierced at C with a small hole, so that the upper surfaces of the mercury in the vertical tubes are at the same level as at C.

The pressures at A and B being equal, and those at G and H also equal, it follows that a column of mercury equal in height to the vertical distance (h) between H and G represents the difference in the pressures due to two columns of mercury, each of height AA′ or BB′, one hot and the other cold. Hence, if the temperatures of the cold column and of the columns DG and FH are the same, the effective height of the cold column is BB′ less the height GH, while the height of the hot column is BB′ or H. Substituting these values in the expression for a obtained above, we get

$$a = \frac{1}{t}\left(\frac{h}{H - h}\right) \quad . \quad . \quad . \quad . \quad (53)$$

where t is the temperature of the hot column, and the temperature of the cold column is taken to be zero. The mean obtained for the coefficient of cubical expansion of mercury between 0° and 100° is 0·0001819.

A knowledge of the absolute expansion of mercury is required to correct the readings of a barometer to allow for the effect of temperature. Since the pressure corresponding to a given barometric height depends

on the density of the mercury, if the temperature of the mercury in a barometer is not 0° C. a correction to the observed height must be applied to get the true height. Further, a correction will have to be applied to allow for the effect of the expansion of the scale which is used to measure the height of the column, for it is only at one temperature that the divisions of the scale are correct. We may consider these two corrections together.

Reduction of the height of barometer to 0° C.

If a is the coefficient of linear expansion of the scale, and the scale is correct at 0° C., the actual length of the portion of the scale corresponding to the observed reading h_t, when the temperature is t, is $h_t(1+at)$.

Since the height of a column of liquid which will exert a given pressure varies inversely as the density of the liquid, if D_0 is the density of mercury at 0° and D_t the density at t, we have

$$\frac{\text{Height of mercury column at } 0^\circ}{\text{Height of mercury column at } t}=\frac{D_t}{D_0}$$

But if v_0 and v_t are the volumes of unit mass of mercury at 0° and t° respectively, and b is the coefficient of cubical expansion of mercury,

$$v_t=v_0(1+bt)$$

and
$$\frac{D_t}{D_0}=\frac{v_0}{v_t}=\frac{1}{1+bt}$$

Hence if H is the height the barometric column would have if the mercury and scale were both at 0°, we have

$$H=\frac{1+at}{1+bt}h_t$$

Since both a and b are small, this equation may be written

$$H=(1+at)(1-bt)h_t=\{1-(b-a)t\}h_t$$

If the scale is of brass, $a=\cdot000020$ and for mercury $b=\cdot000182$ (the temperature being measured in degrees centigrade). Hence for a barometer with a brass scale, we have

$$H=(1-0\cdot000162t)h_t$$

by means of which equation the observed height h_t can be reduced to the value it would possess if the temperature were zero.

62. Density of Water at different Temperatures. Point of Maximum Density.—The expansion of water with rise of temperature is anomalous, since this substance has a maximum density at a temperature of 4° C., so that water when cooled below 4° *expands*.

This property possessed by water has an important bearing in nature, for otherwise all deep lakes in temperate zones would become frozen into a solid block of ice, and only the upper surface would thaw in the summer. As it is, in winter the surface water becomes cooled by radiation, &c., and as it cools, and its temperature falls, it becomes denser and sinks, convection currents being set up. This convection goes on till the temperature of the whole mass of water has fallen to 4° C. On the surface water becoming further cooled it expands, and its density becomes less than that of the water beneath. Hence the colder water remains on the top, and convection currents are not set up. Water being a very bad conductor of heat (§ 77), it takes a long time for the deeper layers of water to part with their heat, and so, even in the hardest winters, the ice in temperate zones is seldom very thick, and the water at the bottom of deep lakes is never colder than 4°.

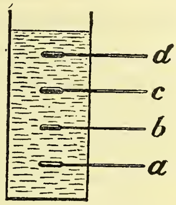


FIG. 98.

The experiments made by Despretz in order to determine the temperature of maximum density illustrate very clearly the changes in temperature which take place in a mass of water, such as a lake. The apparatus, which is a modification of one devised by Hope, consists of a tall metal cylinder fitted with a lid, through the side of which are inserted four thermometers (Fig. 98). The cylinder was filled with water at a temperature of about 10° C., and was then placed in a cold

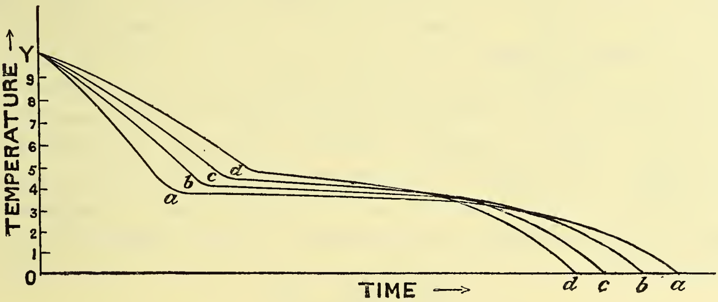


FIG. 99.

room, of which the temperature was about 0°. The temperatures of the water at different depths, as indicated by the thermometers, were noted at frequent intervals, and the results plotted on curve paper, the ordinates being the temperature, and the abscissæ the times. In this way the curves shown in Fig. 99 were obtained. As the water near the sides lost its heat by radiation it became cooler and sank. Thus thermo-

meter a sank the most rapidly, till it indicated a temperature of 4°. Next thermometers b, c, and d arrive at the same temperature in succession. The whole mass being at the point of maximum density, when the temperature fell any more, the colder water was lighter, and hence the upper thermometer (d) began to fall first, the others following in the order c, b, a. These changes are well indicated by the curves. The reason that the horizontal parts of all the curves are not coincident, is that disturbing currents are produced by the manner in which the water is cooled, namely, from the sides of the vessel.

The presence of common salt (sodium chloride) dissolved in water causes the point of maximum density to occur at a lower temperature, and as we shall see later (§ 73) it also lowers the freezing-point. The lowering of the point of maximum density is, however, more rapid than the lowering of the freezing-point, so that for a solution containing about 2·4 per cent. of salt, the point of maximum density occurs at the freezing-point $-1{\cdot}5°$ C. Since sea-water contains about 2·7 per cent. of dissolved salt, there is no point of maximum density above the freezing-point, and hence a mass of sea-water, unlike fresh water, can freeze from the bottom.

63. Expansion of Gases.—In the case of solids and liquids we have not considered the pressure when dealing with the coefficient of expansion, since for moderate changes of pressure the volume is not appreciably affected. In the case of gases, however, it is necessary to take account of changes of pressure, and when considering the effect of a change in temperature we may either measure the change in volume when the pressure remains constant, or the change in pressure when the volume is kept constant.

We have (in § 59) defined the temperature scale in terms of the change in pressure of hydrogen when the volume is kept constant, that is, we have taken the change in pressure in the case of this gas is proportional to the change in temperature. Thus if P_0, P_t, and P_{100} are the pressures to which a given mass of hydrogen has to be subjected to keep the volume constant at the temperatures of 0°, t, and 100° C. respectively, we have

$$100 = x(P_{100} - P_0)$$

where x is a constant factor.

Also
$$t = x(P_t - P_0)$$

Hence
$$t = (P_t - P_0)\frac{100}{P_{100} - P_0}$$

or
$$P_t = P_0\left(1 + \frac{P_{100} - P_0}{100P_0}t\right) \quad . \quad . \quad . \quad (54)$$

But $\frac{P_{100}-P_0}{100P_0}$ is the coefficient of increase of pressure per 1° C., for it is the increase of pressure divided by the original pressure and by the temperature interval, just as the coefficient of expansion is the increase in volume divided by the original volume and by the temperature interval. Hence, calling the coefficient of increase in pressure a_p, we get

$$P_t = P_0(1 + a_p t) \quad . \quad . \quad . \quad . \quad (55)$$

an expression of the same form as that already obtained in the case of the ordinary coefficient of expansion.

If now the volume of the hydrogen considered above is V_0 under a pressure of P_0 at 0° and V_t under a pressure P_0 at t, and a_v is the coefficient of expansion, *i.e.* the coefficient of increase of *volume* at *constant pressure*, we have

$$V_t = V_0(1 + a_v t) \quad . \quad . \quad . \quad . \quad (56)$$

Multiplying equation (55) by V_0 and equation (56) by P_0 we have

$$P_t V_0 = P_0 V_0(1 + a_p t)$$

$$P_0 V_t = P_0 V_0(1 + a_v t)$$

But P_t and V_0 are the pressure and volume of the gas at a temperature t, and P_0 and V_t are also the pressure and volume of the same mass of gas at a temperature t. Hence, by Boyle's law, $P_t V_0 = P_0 V_t$.

Therefore $$P_0 V_0(1 + a_p t) = P_0 V_0(1 + a_v t)$$

or $$a_p = a_v \quad . \quad . \quad . \quad . \quad . \quad (57)$$

This shows that in the case of a gas which obeys Boyle's law the two coefficients are the same. Experiment has confirmed this result, and has shown that as long as the pressure is not much above atmospheric the coefficients have the same value for all gases, a result known as *Charles's law*. The value of the coefficient for the centigrade degree is 0·00366.

Charles's law.

If the temperature of a given mass of a gas is reduced to a temperature $-t$, the volume being kept constant, the pressure P' is given by

$$P' = P_0(1 - at)$$

where $$a = \cdot 00366.$$

If now the cooling be continued down to a temperature of $\left(-\frac{1}{a}\right)^{\circ}$, then

$$P' = P_0(1 - 1) = 0.$$

That is, at this temperature the gas would exert no pressure on the walls of the containing vessel. Now we have reason to believe that a gas at

ordinary temperatures consists of a large number of molecules which are endowed with a continual translatory movement, and that the pressure which a gas exerts on the walls of a containing vessel is due to the bombardment of the walls by these moving molecules. Hence we conclude that when no pressure is exerted the molecules must have come to rest. Further, there is evidence to show that the motion of the molecules depends on the temperature. If, then, we suppose that the lowest possible temperature imaginable is one at which the molecules have entirely lost their motion of translation, we conclude that this zero of temperature is $-\left(\frac{1}{a}\right)^{\circ}$ C. This temperature is called the *absolute zero of the gas thermometer*, and is equal to $-\frac{1}{\cdot 00366}$ or -273° C. If T is the temperature measured from the absolute zero[1] corresponding to t° C., then $T=t+273$, and equation (55) gives

Absolute zero of temperature.

$$P_t=P_0\left(1+\frac{t}{273}\right)=\frac{P_0(273+t)}{273}=\frac{P_0T}{T_0}$$

where T_0 is the temperature of melting ice on the absolute scale. Hence

$$\frac{P_t}{P_0}=\frac{T}{T_0}$$

Similarly, if $P_{t'}$ is the pressure for the same mass of gas, at constant volume, at an absolute temperature T',

$$\frac{P_{t'}}{P_0}=\frac{T'}{T_0}$$

Hence

$$\frac{P_t}{P_{t'}}=\frac{T}{T'} \quad . \quad . \quad . \quad . \quad (58a)$$

That is, at constant volume the pressure is proportional to the absolute temperature.

In the same way it can be shown that

$$\frac{V_t}{V_{t'}}=\frac{T}{T'} \quad . \quad . \quad . \quad . \quad (58b)$$

That is, at constant pressure the volume is proportional to the absolute temperature.

We have now to consider what connection exists between the pressure, volume, and temperature of a gas when all three vary at the same time. Let p_0v_0 be the pressure and volume of unit mass of the gas at 0° C., and

[1] Temperatures measured from the absolute zero of the gas thermometer will be indicated by a capital T.

pv be the pressure and volume at a temperature T on the absolute scale. If the temperature were kept at 0° C. [T_0] and the pressure changed from p_0 to p, by Boyle's law the volume v' would be such that $p_0v_0=pv'$ or $v'=p_0v_0/p$. If now the temperature of a volume v' of gas at a constant pressure p is increased from T_0 to T, we have

$$\frac{v}{v'}=\frac{T}{T_0}$$

Hence, substituting for the value of v',

$$V=\frac{p_0v_0T}{pT_0}$$

or

$$vp=\frac{p_0v_0}{T_0}T \quad . \quad . \quad . \quad . \quad (60)$$

Now in this equation, if p_0 is the standard atmosphere, v_0 is the volume of unit mass of the gas at this standard pressure and at the standard temperature 0° C., so that for any *given* gas p_0v_0/T_0 is a *constant* quantity which we may call R. Hence

The gas equation.

$$vp=RT \quad . \quad . \quad . \quad . \quad . \quad (61)$$

This equation is called *the* gas equation, and, when using it, it must be remembered that v is the volume of unit mass of the gas and R has different values for different gases.

According to Avogadro's hypothesis, the number of molecules of gas contained in a given volume at a given pressure and temperature is the same for all gases. Hence if m is the mass of a molecule of one gas and m' that of another gas, and N is the number of molecules contained in unit volume under standard conditions of pressure and temperature, the mass of unit volume of the first gas will be mN and that of the second gas $m'N$. Hence the volume of unit mass (v) of the first gas will be $1/mN$, and that (v') of the second gas $1/m'N$. Thus for the first gas R will be equal to $\frac{p_0}{mNT_0}$, while for the second gas $R'=\frac{p_0}{m'NT_0}$. Equation (61) gives for the two gases

$$\frac{mvp}{T}=\frac{p_0}{NT_0}$$

and

$$\frac{m'vp}{T}=\frac{p_0}{NT_0} \quad . \quad . \quad . \quad . \quad (62)$$

where the right-hand sides are identical. Now mv is the volume occupied by m grams of the gas, where m is the molecular weight. This fact is generally expressed by saying that mv is the volume of

the gram-molecule, and it represents the volume occupied by a number of grams of the gas equal to the molecular weight. Thus if v_m and v'_m are the volumes of the gram-molecule of the two gases, equations (62) give

$$\frac{v_m p}{T} = \frac{v'_m p}{T} = X$$

where X is a constant which has the *same value for all gases*, so that the gas equation takes the form

$$v_m p = XT \quad . \quad . \quad . \quad . \quad (63)$$

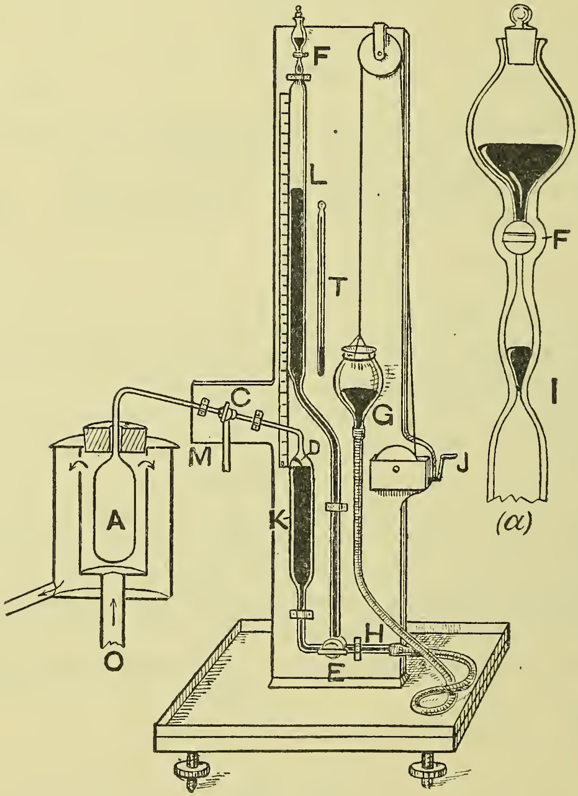


Fig. 100.

64. Measurement of the Thermal Expansion of Gases. The Gas Thermometer.—An instrument for measuring the increase of pressure at constant volume of a gas is shown in Fig. 100. The gas to be examined is contained in the bulb A, which is connected by a fine

capillary tube C with the vessel K. The lower part of K communicates with an upright glass tube L which is closed at the top and with a movable reservoir G. Mercury fills K, L, and G as shown, the upper part of L being free from air, and forming a Torricellian vacuum. When the bulb A is surrounded with melting ice, the height of the reservoir G is adjusted so that the mercury in K just touches the end of a black glass point D, and the height of the mercury surface in L above this point is read off on the scale placed alongside L. This height gives the pressure to which the gas is subject. It will be noticed that the column of mercury of height LD acts both as barometer and manometer, and hence gives directly the total pressure to which the gas in A is subjected. Steam is then passed through the vessels surrounding A, and the pressure is again adjusted so that the mercury surface coincides with the point D, and the new height of the mercury column LD gives the pressure required to bring the gas back to its former volume when the temperature is 100°. Knowing in this way the pressure at 0° and at 100°, an approximate value of the coefficient of expansion can be calculated. This approximate value will require correction to allow for the effect of the increase in size of the bulb as the temperature is raised, which can be calculated if the coefficient of expansion of the material of the bulb is known. Another correction is necessitated by the fact that a certain amount of the gas is contained in the tube C and in the space round the pointer D, and that the temperature of this portion of the gas remains practically constant throughout the experiment, but owing to the increase of pressure a larger proportion of the gas is contained in this "dead space" at the higher temperature than at the lower. If a large range of temperature is employed or very great accuracy is aimed at, a further correction has to be applied to allow for the expansion of the bulb due to the increase in the internal pressure.

Constant-volume air thermometer.

A piece of apparatus for measuring the coefficient of increase in volume at constant pressure is shown in Fig. 101. The bulb A containing the gas is connected by a capillary tube with a reservoir B which is surrounded by water, so that its temperature may remain constant. A side branch of the capillary is connected to a manometer C containing oil. When the bulb A is surrounded by melting ice the mercury in B is brought up to a mark K, the tap F being turned so that the pressure in A is the atmospheric pressure. This tap being closed and the tap E so turned as to cut off the connection between B and G, the bulb A is surrounded by steam. When the temperature has become constant, mercury from B is run off through the tap D into a weighed beaker till the manometer C indicates that the pressure is the same inside and out. The mercury which has been

Constant-pressure air thermometer.

drawn off is then weighed, and from this weight we can calculate its volume, and this volume represents the volume by which the gas in A has expanded, the principle of the method being similar to that used in the weight thermometer. The volume of the bulb A up to the point in the capillary where the heating extends has to be determined. This is obtained by weighing the mercury required to fill the bulb up to this point. As in the previous case, a correction has to be applied to allow for the expansion of the bulb.

Either of the arrangements described above may be used as a gas

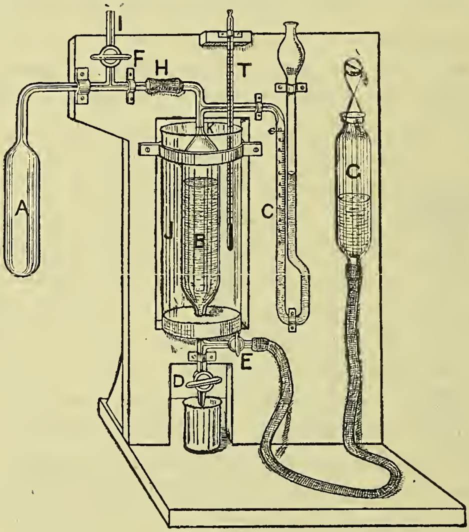


FIG. 101.

thermometer, for if P_0 and P_{100} are the pressures, at constant volume, corresponding to the temperatures of melting ice and boiling water, and P_t is the pressure corresponding to the unknown temperature t, then, as on p. 140, we have

$$t=(P_t-P_0)\frac{100}{P_{100}-P_0}$$

with a similar expression in the case of change of volume at constant pressure.

CHAPTER II

CALORIMETRY

65. Quantity of Heat. Specific Heat.—We have in the preceding chapter considered temperature and some of the changes which are due to changes of temperature. We have now to consider how the temperature of a body will alter when heat is communicated to or abstracted from it. If we have two vessels containing water at 100° C., but one containing twice as much water as the other, it is a familiar fact that it will take much longer for the larger vessel to cool than for the smaller. Further, to heat up the larger vessel through a given temperature range over a given flame will take longer than to heat the smaller vessel through the same range of temperature. To account for the above effects, we may say that the *quantity of heat* required to raise the temperature of the larger mass of water by a given amount is greater than that required to raise the temperature of the smaller quantity by the same amount.

As a unit of quantity of heat we may take the amount of heat which has to be communicated to unit mass of water to raise its temperature through one degree. There will be different units according as to what units of mass and temperature we employ. The heat required to raise the temperature of one gram of water through 1° C. is almost exclusively used in scientific measurements, and is called the *calorie*. If the unit of mass is a kilogram, the unit of heat is called a large calorie, and is equal to 1000 calories. The heat required to raise a pound of water through 1° F. is often used as a unit in engineering, and is called a *British Thermal Unit*, being indicated by the letters B.Th.U.

Units of quantity of heat.

As we shall see later, it is possible to measure quantity of heat in work units, and, particularly in electrical engineering, a unit often employed is $4{\cdot}2 \times 10^7$ ergs, this unit being called a *joule*.

The following table shows the connection between the above units:—

1 joule = 0·2387 calorie.
1 B.Th.U. = 252 calories.
1 joule = 0·000948 B.Th.U.

Careful measurement have shown that the quantity of heat required to raise the temperature of a given mass of water through one degree varies slightly at different temperatures. Thus taking as unity the heat required to raise the temperature of one gram of water from 19·5° C. to 20·5° C., the heat required to raise the temperature of one gram of water through one degree at 0° is 1·0094, at 40° is ·9982, and at 100° is 1·0074. Hence in accurate work it is necessary to specify the temperature at which the calorie is defined.

Specific heat.

The quantity of heat required to raise the temperature of unit mass of any substance through one degree is called the *specific heat* of the substance. The value of the specific heat for any given substance will not depend on the units of mass and of temperature employed, since these will also be involved in the definition of the unit of heat; in fact we may define the specific heat of a substance as the ratio of the quantity of heat required to raise the temperature of a given mass of the substance through a given temperature range to the quantity of heat required to raise the temperature of an equal mass of water through the same range of temperature.

66. Measurement of Specific Heat.—The most usual method of determining the specific heat of a solid is called the method of mixtures, and consists in heating a given mass of the solid to a known temperature, and then immersing it in a vessel containing a known mass of water, the initial and final temperatures of which are noted. If M is the mass of the body, W that of the water, t_1 the initial temperature of the body, t_2 the initial temperature of the water, and t_3 the final temperature of the body and the water, we have, if we suppose for the moment that the vessel does not take up any heat, that the heat gained by the water is $W(t_3-t_2)$. The heat lost by the body is $M(t_1-t_3)s$, where s is the specific heat of the body. Equating these two quantities of heat, we get

$$s=\frac{W(t_3-t_2)}{M(t_1-t_3)} \quad . \quad . \quad . \quad . \quad (64)$$

Method of mixtures.

Since the temperature of the vessel in which the water is contained (called the calorimeter) is raised from t_2 to t_3, some of the heat given out by the body will have been used for this purpose, and the above result must be corrected on this account. If w is the mass of the calorimeter, and c the specific heat of the material of which it is composed, the heat necessary to raise its temperature from t_2 to t_3 is $w(t_3-t_2)c$. The product wc, which represents the quantity of water which would require the same quantity of heat to raise its temperature one degree as does the calorimeter, is called the *water*

equivalent or *water value* of the calorimeter. The heat gained by the water and calorimeter is $W(t_3-t_2)+w(t_3-t_2)c$, and hence

$$s=\frac{(W+wc)(t_3-t_2)}{M(t_1-t_3)} \quad . \quad . \quad . \quad . \quad (65)$$

In forming the above expressions, we have supposed that *all* the heat given out by the hot body is received by the calorimeter and its contents. Since the hot body has to be moved from the enclosure in which it was heated to the calorimeter, special precautions have to be taken to prevent loss of heat during transit. Again, although the calorimeter may be at the same temperature as its surroundings at one temperature, say the initial temperature, yet when the hot body is placed in, the temperature will be higher than that of the surroundings, and hence the calorimeter will lose heat by conduction and radiation. In order to reduce such loss of heat to a minimum, the calorimeter is supported on small, badly conducting feet, or suspended by threads, so that it shall not gain or lose heat by conduction through the supports. The loss or gain of heat by radiation is reduced to a minimum by having the outside of the calorimeter polished, since polished metal is a bad radiator.

Rumford first proposed to eliminate the effects of radiation by making a preliminary experiment to determine approximately the rise in temperature of the calorimeter, and then, in the final experiment, cooling the calorimeter before the introduction of the hot body to a temperature below that of the surrounding bodies by an amount equal to half the rise. By this device, during the first part of the time between the introduction of the hot body and the reading of the final temperature, the calorimeter would receive heat by radiation, and during the second part it would lose heat. As, however, the temperature rises most rapidly at first, this correction is not perfect, and for very accurate observations the loss of heat during the interval which elapses between the introduction of the hot body and the instant when the maximum temperature is reached has to be actually determined, and an allowance made on this account.

The method of mixtures is also applicable to the determination of the specific heat of liquids, for if the water in the calorimeter is replaced by the liquid, and we either know the specific heat of the solid, or determine this quantity by a preliminary experiment, then from the rise in temperature which occurs when the solid is introduced we can immediately deduce the specific heat of the liquid.

A form of calorimeter suitable for determining the specific heats of solids or liquids by the method of mixtures is shown in Fig. 102. The substance of which the specific heat is to be measured is heated in the

heater A. This heater is shown in section at (*a*), and is connected by the side-tube E with a boiler, so that steam enters at E and passes out through F to a condenser. The temperature to which the substance is heated is indicated by the thermometer T_1. The calorimeter C is suspended by means of three light strings inside a brightly polished metal vessel D, while this vessel is itself contained within a wooden box B (one side is shown as removed in the figure to exhibit the calorimeter). A delicate thermometer T_2, which is held in a clip attached to the box B, is used to give the temperature of the liquid in the calorimeter, while a stirrer S serves to mix the liquid and thus ensure it all being at the same temperature. A screen K, which slides up and down in guides, serves to protect the calorimeter from radiation from the heater. When the substance has attained the temperature of the heater, the screen K is raised, the box B run on its guides under the heater, and the substance dropped down into the calorimeter, the small slide L (Fig. 102 (*a*)) being momentarily drawn out for this purpose. Directly the substance has been introduced, the calorimeter is withdrawn, and the screen K again lowered.

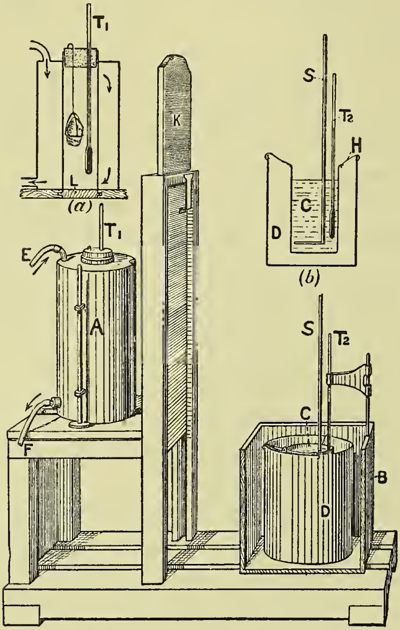


Fig. 102.

The consideration of calorimetric methods which depend on latent heat of vaporisation or fusion will be dealt with later (§ 72).

In the case of gases we have to distinguish two cases, first the specific heat when the pressure is kept constant, and secondly the specific heat when the volume is kept constant. When the pressure is kept constant the gas will increase in volume when the temperature rises, and hence will have to drive back the atmospheric pressure, that is, to do external work. Now, as we shall see later, this work is performed at the expense of the heat communicated to the gas, so that more heat has to be supplied when the volume increases than when the volume remains constant. In the case of liquids and solids the change in volume with temperature is so small that the amount

Two specific heats in the case of gases.

of heat required to do the external work is quite negligible. With gases, however, this is by no means the case, so that the heat required to raise the temperature of a given mass of gas at constant pressure, when it expands and does external work, is considerably greater than that required when the volume is kept constant, so that no external work is performed.

Determination of the specific heat of a gas at constant pressure.

One of the most accurate methods of measuring the specific heat of a gas at constant pressure is one devised by Callendar, and called by him the continuous-flow electrical method. The principle of the method is to pass a current of gas through a tube and determine the rise in temperature when heat is supplied to the gas at a known rate. The arrangement used in the experiment is shown in Fig. 103. A steady stream of the gas is forced through the tube A, and its temperature is measured by the platinum thermometer T_1 (see § 153). After traversing the spiral B, the gas passes a coil of wire H through which a measured electric current is

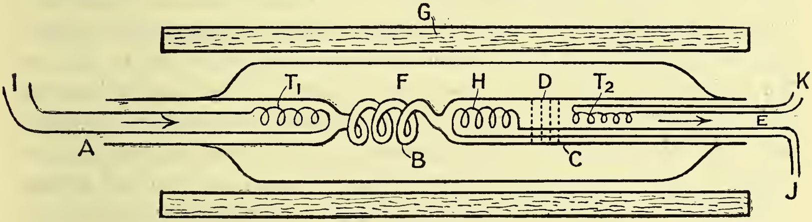


FIG. 103.

passed. Hence, knowing the resistance of this coil, the heat, H, communicated to the gas per second can be calculated. After passing through some wire-gauze screens D, the object of which is to make the temperature of the gas uniform throughout, the gas passes a second platinum thermometer T_2 by means of which the temperature of the gas after it has received the heat is determined. The central tube is surrounded by a large glass tube F, and the intervening space is evacuated, since (§ 77) in this way the loss of heat can in a great measure be prevented. A water-jacket G serves to protect the instrument from variations in temperature due to outside sources. If the difference in temperature of the gas before and after receiving the heat is t, and the mass of gas passing per second is m, the specific heat is, to a first approximation, given by

$$S = \frac{H}{mt}$$

for after the gas has been passing for some time the temperature of every point of the tube through which it passes remains constant, so that all

the heat supplied is employed in heating the gas. A small correction has to be applied for heat lost by the gas between the coils H and T_2, the value of the correction being obtained by performing experiments with different rates of flow of the gas.

The direct measurement of the specific heat at constant volume is described in § 72.

67. Dulong and Petit's Law. Variation in the Specific Heat with Temperature.—The product of the specific heat of an element in the solid state into the atomic weight is called the atomic heat, and is nearly constant for all elements, and equal to 6·4. The product of the specific heat into the atomic weight of all gases is also approximately equal to a constant, this constant being about 3·4. This law, known as Dulong and Petit's law, has various exceptions; thus the specific heat of carbon in the form of diamond at ordinary temperatures is 0·144, and hence the product of the atomic weight (12) into the specific heat is equal to 1·7. It is found, however, that the specific heat of diamond increases quite rapidly with the temperature, so that for a temperature of 1000° C. it has the value 0·459, and thus the atomic heat at this temperature is 5·5, a value much more nearly equal to that for the majority of elements. Two other elements, iron and silicon, for which the atomic heat is very low at ordinary temperatures, resemble carbon in that the specific heat increases rapidly with temperature, and hence at high temperatures these elements more nearly obey Dulong and Petit's law. Elements which at ordinary temperatures obey the law fairly well, in some cases depart widely at very low temperatures. Thus the atomic heat of aluminium at about 100° C. is 6·0, while the value at – 170° C. is 3·2.

In the case of a substance which can exist in more than one allotropic modification there is generally a marked difference in the specific heats of the different forms. Thus at ordinary temperatures the specific heat of diamond is 0·147, of wood charcoal 0·242, and of graphite 0·202.

The specific heat of most substances is different in the three states—solid, liquid, and gas. Thus the specific heat of ice is 0·5, and that of steam 0·477; that of water of course being unity. The probable cause of these differences is that when the temperature of a substance is raised some work has to be done in separating the molecules against their mutual attractions, and hence as these attractions vary according to the state, the heat required is also different.

CHAPTER III

CHANGE OF STATE

68. Melting-Point.—The temperature at which the solid and liquid forms of a substance can coexist without change is called the melting-point of the substance. Thus if we have a vessel containing a mixture of ice and water, if the temperature is below 0° C. the water will gradually change into ice, while if the temperature is above 0° C. the ice will gradually melt. At the melting-point (0° C.), however, the proportion of ice to water will remain constant. In the case of ice and water the change from the solid to the liquid state is quite abrupt, and the melting-point is well defined. Some substances, such as glass, pass gradually from the solid to the liquid state, the solid getting gradually softer and softer as the temperature rises, and in such a case the melting-point is very ill defined.

Change of volume on melting.

For the majority of substances there is an increase of volume on melting, so that the density of the solid is greater than that of the liquid, and hence at the melting-point the unmelted solid sinks in the liquid. There are, however, exceptions to the above; some substances, notably ice, cast iron, and bismuth, contract on melting.

In the case of water the density of ice at 0° C. is 0·91674, and that of water at the same temperature is 0·99987, so that the increase in volume of a gram of water as it solidifies is 0·0907 c.c. This behaviour of water is of great importance in nature, for as it is, ice *floats* on the surface of water, and hence when the temperature of the air rises the ice at once melts. If, however, ice were more dense than water the ice would sink to the bottom, so that the heat necessary to melt the ice would have to be transmitted through the water, and hence during a long winter the ice would gradually accumulate, and lakes and rivers, even in temperate zones, would probably become entirely filled with ice.

Water, when changing to ice, is capable of exerting an enormous force if its expansion is resisted. This expansive force is evident in the bursting of water pipes and the disintegration of rocks into the pores of which water has penetrated, produced when the water has frozen.

It can be shown, as was first done by James Thomson, that if a body *expands on solidification*, as is the case with water, then increase of

pressure will *lower* the melting-point, while if the body *contracts on solidification* increase in pressure will raise the melting-point. An increase of pressure of one atmosphere lowers the melting-point of water by 0·0075° C., so that to keep water liquid at a temperature of −7·5° C. a pressure of a thousand atmospheres must be exerted. Thus if water fills a closed vessel and the temperature is −7·5° it is capable of exerting a pressure of a thousand atmospheres.

Melting-point of ice lowered by increase of pressure.

The lowering of the melting-point of ice with pressure is the reason why it is possible to form a snow-ball. Owing to the force exerted by the hands the ice at the points where the snow crystals touch is compressed sufficiently to cause local melting, and on the release of the pressure the water produced, which is at a lower temperature than 0° C., again freezes, thus binding the crystals together. If the temperature of the snow is much below 0° C., the pressure which can be exerted is not sufficient to cause the melting-point to fall to this low temperature, and hence the snow refuses to "bind."

69. Vapour Pressure of Liquids. Boiling-Point.—If a small bubble of air is allowed to pass into the Torricellian vacuum of a barometer, the mercury column is depressed, and if a succession of bubbles are passed up, each will produce a depression. If a small drop of a liquid, say ether, is introduced the column will be depressed and the ether become entirely vaporised. If, however, successive small drops of ether are introduced, it will be found that after a time the further addition of ether does not produce an *additional depression*, and that the ether no longer vaporises, but simply floats as a liquid on the top of the mercury column. If the space above the mercury be increased or decreased, by raising or lowering the barometer-tube in the cistern, it will be found that, so long as there is any liquid present, the *height* of the mercury column remains constant, but that the quantity of ether which vaporises varies with the space above the mercury. If the temperature is increased, more ether vaporises, and the mercury column becomes more depressed, and *vice versâ*.

The depression of the mercury column indicates that the liquid forms a vapour in the Torricellian vacuum, and that this vapour exerts a pressure on the upper end of the mercury column which partly balances the atmospheric pressure. The amount by which the column is depressed is a measure of this pressure, which is called the *vapour pressure* of the liquid. When an excess of liquid is present, so that the vapour exerts its maximum pressure, and no more liquid will vaporise at the given temperature into the space under consideration, the vapour is said to be *saturated*. When, however, a given space contains some vapour, but if some more liquid were introduced some of it would vaporise

at the given temperature, the vapour is said to be *unsaturated*, or *super-heated*.

The vapour pressure, or tension as it is sometimes called, of a liquid depends on the temperature only. In the case of non-saturated vapours, Boyle's and Charles's laws apply approximately, the approximation being the better the further the vapour is removed from its saturation-point.

Suppose some liquid is contained within a cylinder which is closed by a freely moving piston, and that a pressure P acts on the outside of this piston. If we start with the temperature of the liquid sufficiently low, the vapour tension will be less than P, and the pressure of the vapour on the inside of the piston will be less than that on the outside, so that the piston will rest on the surface of the liquid. As the temperature of the liquid is raised, the vapour pressure increases; and when the vapour pressure is equal to the pressure P acting on the outside of the piston, this latter is in equilibrium. If the temperature rises ever so little more, the vapour pressure will be greater than P, and so the piston will be driven out, and vapour will be formed freely above the liquid. Now, exactly the same thing occurs when a liquid is heated in an open vessel, so that, when vapour is formed freely, the vapour pressure is equal to the pressure of the atmosphere acting on the surface of the liquid. Since when a liquid is vaporising freely it is said to *boil*, we have that when a liquid boils the vapour pressure at that temperature is equal to the external pressure to which the liquid is subject, while at temperatures below the boiling-point the vapour pressure is less than the external pressure. At temperatures above the boiling-point, corresponding to the pressure acting, the *liquid* cannot exist, and the vapour is unsaturated. We may define the boiling-point of a liquid under a given pressure as that temperature at which both the liquid and the saturated vapour can coexist. The term *the* boiling-point is, however, generally reserved for the case where the pressure is one standard atmosphere.

Boiling-point of a liquid.

The curve given in Fig. 104 shows the vapour pressure of water at different temperatures. This curve divides the diagram into two regions, in one of which the conditions are such that the water can only exist as an unsaturated vapour, and in the other only as a liquid, while along the curve, which is called the *steam line*, we may have the liquid and vapour existing simultaneously, *i.e.* the vapour is saturated. For suppose we had some water enclosed in the Torricellian vacuum of a barometer-tube, the pressure being 20 cm. of mercury and the temperature 80°. The conditions are represented by the point A on the diagram. From the curve we see that the maximum vapour pressure corresponding to a temperature of 80° is 35·5 cm., so that the vapour is unsaturated. If now the pressure

The steam line.

on the vapour is increased, the temperature remaining constant, the conditions the vapour passes through are represented by the vertical line AB. When the point B on the curve is reached, the pressure is equal to the maximum vapour pressure, and if the pressure is increased beyond this point the vapour will condense into a liquid. In the same way if, starting from A, we keep the pressure constant, reducing the temperature, the changes are indicated by the straight line AC. When the point C is reached, *i.e.* the temperature falls to 66°, the vapour will be saturated. Any further fall of temperature will be accompanied by the condensation of the vapour into a liquid. Hence, corresponding to all points on the diagram to the right and below the curve we have vapour, and to those on the left and above we have liquid.

In addition to the curve considered above which gives the relation

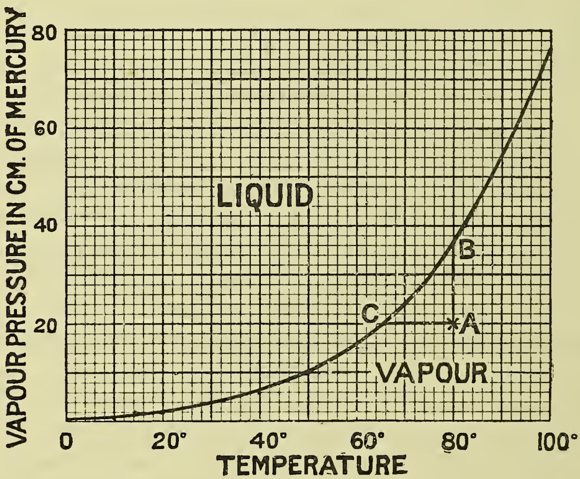


FIG. 104.

between the pressure and temperature when water and steam can coexist, we may draw a similar curve, called the *ice line*, showing the relation between pressure and temperature when *ice* and *water* can coexist. This curve will of course give the melting-point of ice for different pressures. It will pass through the point corresponding to a pressure of one atmosphere and a temperature of 0° C., and since increase of pressure causes a lowering of the melting-point, it must slope upwards and to the left. The lowering of the melting-point with pressure, however, is so small that on the scale of Fig. 105 the ice line is practically a vertical straight line. All points to the left of the ice line correspond to ice, while points to the right of the ice line and the left of the steam line correspond to liquid water. The

The ice line.

ice line cuts the steam line at a pressure of 0·46 cm. of mercury and a temperature a very little above 0° C.

At temperatures below 0° and at sufficiently low pressures it is found that ice can pass directly into the condition of a vapour, *i.e.* steam, without passing through the intermediate liquid state, or conversely vapour will condense in the form of solid. A familiar example of this change is the production of hoar-frost, when the moisture in the air is deposited as ice, without the intermediate formation of rain or dew. The direct passage of a solid into a vapour is called *sublimation.* Camphor is a familiar example of a substance which sublimes at ordinary pressures. In order to *melt* camphor it must be heated at a pressure higher than atmospheric.

A curve showing the relation between the pressure and temperature

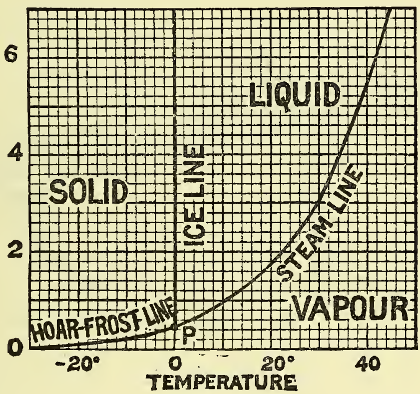


FIG. 105.

when ice and water vapour can coexist is called *the hoar-frost line.* The hoar-frost line passes through the point P (Fig. 105), where the ice line and the steam line intersect. Hence at the pressure and temperature corresponding to the point P we can have ice, water, and water vapour coexisting unchanged. For this reason P is called the *triple point.*

Sublimation. The hoar-frost line.

The vapour pressure of liquids at ordinary temperatures and pressures less than atmospheric can be determined by introducing a little of the liquid above the mercury in a barometer and noting the depression produced. This method, however, has only a very limited application, and Regnault devised an arrangement by which the temperature was automatically kept constant during an experiment, and which could be used to measure the vapour pressure at all temperatures. The liquid of which the vapour tension

Measurement of vapour pressure.

is to be measured is contained in an air-tight metal vessel A (Fig. 106), from which an inclined tube leads to a reservoir, B. This reservoir is surrounded by water so that its temperature may remain constant, and can be connected by the three-way cock F either to a compressing or exhausting pump or to a manometer. The liquid in A is heated, and when it boils the vapour passes up into the inclined tube, where it is condensed by a stream of cold water which passes through the condenser C, and then flows back into the boiler A. The temperature of the vapour over the boiling liquid is given by four thermometers T, which are placed in four iron tubulars, which are closed at the bottom, and contain mercury. Since a liquid boils when its temperature is such that the maximum vapour pressure is equal to the pressure to which it is subjected, the manometer gives the vapour pressure corresponding

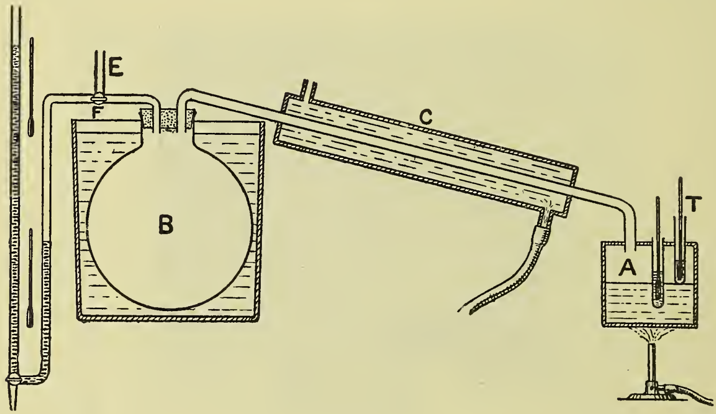


FIG. 106.

to the temperature as given by the thermometers T. The pressure in the globe having been adjusted to the required value, which may if necessary be either greater or less than the atmospheric pressure, the flask is heated, boiling soon starts, and in a very short time the temperature becomes absolutely constant, and remains so as long as is required. The manometer and the thermometers T having been read, the pressure is altered by means of the pump, and when ebullition has continued for a few minutes, the readings for the new pressure are taken, and so on. In this arrangement, when the steady state has been reached, the heat supplied by the burner is employed in supplying the latent heat of vaporisation of the liquid. The vapour then passes to the condenser, where it parts with its latent heat and again becomes liquid, and returns to the vessel A, running down the inclined tube. The pressure does not rise, since as much vapour is condensed during each second as is

produced. If the supply of heat is increased, the rate at which the vapour is produced is also increased, and the only effect of this is that the vapour is able to pass a little further up the condenser before it is all condensed; but since the condenser is always made so long that the vapour never reaches the further end, no vapour passes over to the globe B. Thus the rate at which the vapour is condensed is increased, and remains equal to the rate at which the liquid is vaporised, so that the pressure does not alter.

If we have a vacuous space and introduce more than sufficient liquid to saturate the space, the liquid will at once produce enough vapour to saturate the space, and the pressure will rise to a value h, which is the vapour tension of the liquid at the temperature of the experiment. Next suppose we have the same space filled with a gas at a pressure H, and again introduce the liquid. As before, the liquid will evaporate; the process, however, will be much slower than was the case when no gas filled the vessel, but finally just as much liquid vapour will be produced, and the pressure in the vessel will rise to the value $H+h$.[1] Hence the pressure exerted by, and the quantity of, a vapour which *saturates* a given space are the same for the same temperature, whether this space is filled by a gas or is a vacuum. The pressure exerted by a mixture of gas and a vapour, whether saturated or not, or of two vapours, or of two gases, is equal to the sum of the pressures which each would exert if it alone occupied the given space. The above laws are known as Dalton's laws, since he first enunciated them.

An unsaturated vapour obeys Boyle's and Charles's laws fairly well so long as the temperature is considerably above that at which the vapour would be saturated. Hence a mixture of an unsaturated vapour and a gas also obeys these laws within the same limits.

70. Humidity of the Atmosphere. Hygrometric State.—The quantity of water vapour present in the air varies greatly from time to time, but the effects due to the water vapour present depend more on how nearly the air is saturated than on the absolute quantity of water vapour contained in unit volume of air. If F is the vapour pressure of water at the temperature of the air, then when the air is saturated the partial pressure of the water vapour is F. If, however, the air is not saturated, the partial pressure of the water vapour being f, then the ratio f/F expresses what fraction of the maximum quantity of water vapour which could be taken up by the air is actually present, and is called the *humidity*, or *relative humidity*, of the air.

Humidity.

[1] The quantities H and h, which represent the pressures which would be exerted by the gas and the vapour respectively if either *alone* filled the given space, are called the *partial pressures* of the gas and vapour.

The weight of water vapour in a given volume being proportional to the pressure it exerts, if W is the weight of water contained in a cubic metre of air *saturated* at a given temperature, while w is the weight of water vapour actually present, we have

$$\frac{w}{W}=\frac{f}{F}$$

Thus if the hygrometric state of the air at any instant is 0·5, the air is capable of taking up double the quantity of water that it actually contains.

Owing to the fact that the vapour tension of water increases rapidly with temperature (see Fig. 104), a quantity of moisture which on a cold day would saturate the air, will on a warm day only produce a quite low humidity. Further, a mass of air which is far from saturated, if it becomes cooled down, say when the sun sets, may reach such a temperature that it is saturated, and hence any further cooling will cause a separation of liquid water in the form of dew. The temperature to which the air must be cooled to become saturated is called the *dew-point*. Suppose that air at a temperature t is cooled down to the dew-point t_0, and f is the vapour tension of water at the dew-point, then since by hypothesis the air has been cooled without loss or gain of water, f must also be the tension of the water vapour actually present at the temperature t. Now if we measure the dew-point t_0, the value of f can be obtained from a table giving the vapour pressure of water at different temperatures, while the value of F, which is the vapour tension at a temperature t, can also be obtained from such tables. Thus the hygrometric state can be obtained if we note (1) the temperature of the air t, and (2) the temperature of the dew-point t_0.

The dew-point.

When air expands against pressure, as we shall see later, it does external work driving the pressure back, and in doing so becomes cooled. Hence a mass of air, which is not saturated, if allowed to expand may become so much cooled as to be reduced below the dew-point, so that moisture is deposited. This effect can sometimes be seen in mountainous regions. Warm and unsaturated air is carried by the wind against a mountain side and deflected upwards. As the air rises the pressure falls, and hence expansion occurs, and the air becomes sufficiently cooled to cause the formation of a cloud, which appears to hang round the top of the mountain. After passing the mountain the air is often carried down again in the eddies on the leeward side of the mountain, and as the pressure rises and the air is compressed the temperature rises, and the dew is again dissipated in the form of vapour.

Instruments intended for the measurement of the hygrometric state

of the air are called hygrometers, and may be divided into three classes: (1) Those in which the dew-point is determined; (2) those in which the humidity is deduced from the rate of evaporation of water, called wet and dry bulb hygrometers; and (3) those in which the weight of water vapour contained in a given volume of air is measured by absorbing the moisture and weighing.

Hygrometers.

The most commonly employed dew-point hygrometer is that devised by Regnault. This instrument consists of two glass tubes E and F (Fig. 107), the lower ends of which are closed by thin silver thimbles. They are each closed at the top by a cork, which supports a delicate thermometer (T and t). Through the cork in E a glass tube A also passes, the end reaching nearly to the bottom of the thimble. The tube E is connected by means of a side tube, which fixes it to the stand, and an india-rubber tube C to an aspirator D. Some ether is placed in the thimbles, and after the instrument has had time to reach the temperature of the air, the two thermometers are read, giving the temperature of the air, t. The aspirator is now started, and draws air through the tube A into the instrument. This air bubbling through the ether causes evaporation, which cools the ether and thimble, which in turn cools the air in its immediate vicinity. When a film of dew is deposited on the thimble E, indicating that the dew-point has been reached, the aspirator is stopped, and the temperature of the thermometer T read. It is again read when the dew disappears from the thimble, and the mean of these two readings gives the dew-point t_0. Then by means of tables the vapour tensions at t and t_0 are obtained, and hence the humidity is calculated.

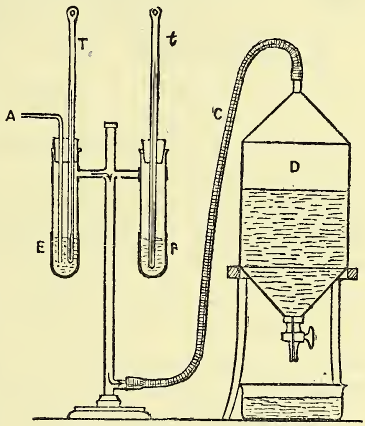


FIG. 107.

The wet and dry bulb hygrometer depends for its action on the fact that the drier the air is, the more rapid will be the evaporation from a wet body exposed to the air. Since evaporation requires the supply of heat (latent heat of evaporation), it follows that the extent to which a wet body is cooled by evaporation will depend on the hygrometric state of the surrounding air. Two similar thermometers are fixed on a stand, the bulb of one of them being covered with muslin kept moist by means of a piece of lamp-wick which dips in a vessel of water. Unless the air is saturated, evaporation will take place from the muslin, and hence the

wet bulb thermometer will indicate a lower temperature than the other, the difference being greater the greater the evaporation, that is, the drier the air. By comparing the readings of the wet and dry bulb thermometers with the humidity, as obtained by other hygrometers, a table has been drawn up, by means of which, from the reading of the dry bulb thermometer, and the difference between the dry and wet bulb thermometers, the dew-point can be obtained. The indications of this instrument are, however, considerably influenced by its environment, also by the action of draughts, hence in some forms of wet and dry bulb hygrometers (Assmann's) a rapid stream of air is drawn past the bulb by means of a fan driven by clockwork. In this way the instrument is always used in approximately the same draught.

71. Latent Heat.—When a vessel containing a mixture of ice and water at 0° is heated, it is found, if the contents are well stirred, that the temperature remains at 0° as long as any ice is left. Since heat is being supplied and the temperature does not rise, it follows that heat must be required to convert ice at 0° into water at the same temperature. This heat, which is employed not in changing the temperature of a body, but in changing its state, is called *latent heat.* In the same way, to convert water at 0° into ice at 0°, heat has to be abstracted. The quantity of heat required to melt unit mass of a solid, or the quantity of heat which must be removed to convert unit mass of a liquid into a solid, in both cases without changing the temperature, is called the *latent heat of fusion* of the substance.

In the same way heat must be supplied to a liquid at the boiling-point to convert it into vapour at the same temperature, and this heat is evolved when the vapour condenses. The quantity of heat which has to be supplied to convert unit mass of the liquid at the boiling-point into vapour at the same temperature is called the latent heat of vaporisation.

In the case of water the latent heat of fusion is 80 calories per gram, and the latent heat of vaporisation is 536 calories per gram.

According to the molecular theory of the constitution of matter, the molecules in a solid are more closely held together by intra-molecular attractions than in a liquid, and part, at any rate, of the latent heat of fusion represents the work which has to be done in loosening the molecules of a solid when it becomes a liquid. A somewhat similar effect occurs when a liquid is converted into a gas, but here, since in general there is a *large* increase in volume, a considerable fraction of the latent heat is employed, as we shall see later, in performing the external work corresponding to the expansion against the atmospheric pressure.

The latent heat of solids may be measured by means of the method

of mixtures. Thus suppose W grams of a solid, of which the latent heat of fusion is L, at a temperature t_1 are placed in a calorimeter, the water equivalent of which and of its contents is w, and that the temperature of the calorimeter falls from t_2 to t_3. If s is the specific heat of the body in the solid state, s^1 the specific heat in the liquid state, and t_0 the melting-point of the body, then the heat absorbed by the body in being heated from t_1 to t_0 in the solid state, then melting at t_0, and finally rising from t_0 to t_3 in the liquid state, is

Measurement of latent heat.

$$Ws(t_0-t_1)+WL+Ws^1(t_3-t_0),$$

while the heat lost by the calorimeter and its contents is

$$w(t_2-t_3).$$

Equating these two quantities of heat, we get

$$L=\frac{w(t_2-t_3)-Ws(t_0-t_1)-Ws^1(t_3-t_0)}{W} \quad . \quad . \quad (66)$$

If the solid is originally at its melting-point, t_1 is equal to t_0, and no heat is used in raising the temperature of the solid, so that L is given by

$$L=\frac{w(t_2-t_3)-Ws^1(t_3-t_0)}{W} \quad . \quad . \quad . \quad (67)$$

In order to determine the latent heat of a vapour, say of steam, steam is passed into a calorimeter containing a weighed quantity of water, and the temperature before and after the passage of the steam is noted. The weight of steam condensed is obtained by again weighing the calorimeter, and the latent heat can be at once calculated in a manner similar to that used above in the case of the solid.

72. Bunsen's Ice Calorimeter. Joly's Steam Calorimeter. — Bunsen has utilised the change in volume, which takes place when ice is melted, to estimate the quantity of ice melted, and hence, knowing the latent heat of ice, to obtain a measure of the heat employed. His ice calorimeter consists of a glass test-tube A (Fig. 108), fused into a cylindrical glass bulb B. The lower part of this bulb is connected by a glass tube C, with a horizontal glass tube D, of fine bore, to which a scale is attached. The upper part of B is filled with distilled water, which has been well boiled to free it from dissolved air, the lower part and

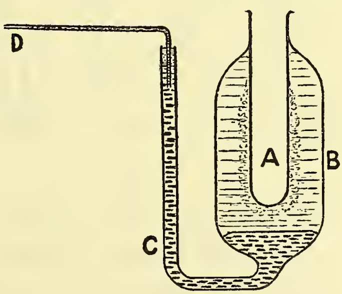


FIG. 108.

the side tube being filled with mercury. By passing alcohol, which has been cooled in a freezing mixture, through A, a coating of ice is formed all round the lower part of A. The instrument is then packed round with melting snow, and left till the temperature of the whole apparatus comes to zero. To determine the specific heat of a substance, a known mass, heated to a temperature t, is dropped into A, and the amount of ice melted calculated from the distance the mercury recedes along the graduated tube. The instrument is often calibrated by introducing a known mass of water at a temperature t, and noting the number of divisions through which the mercury recedes, and then calculating the quantity of heat given to A which causes the mercury to recede through one division.

Another form of calorimeter which depends for its action on latent heat is the Joly steam calorimeter, in which the heat, necessary to raise the temperature of the substance of which the specific heat is being measured from a known temperature of about 20° to 100°, is obtained by determining the weight of steam which must be condensed to supply this heat. The arrangement employed in measuring the specific heat of a gas at constant volume is shown in section in Fig. 109. The gas is contained in a copper sphere A, suspended by means of a fine wire C, from one arm of a delicate balance D, which is supported on a shelf above the apparatus. This wire passes through a small hole in the top of a copper vessel B, which is itself enclosed in a non-conducting box. Steam is admitted to the box B through the tube E, and that which is not condensed within the apparatus passes out through the tube F.

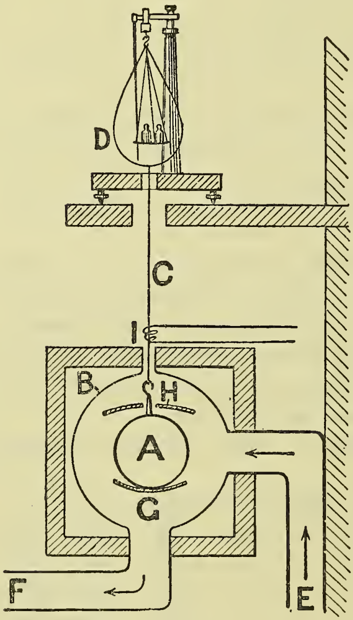


Fig. 109.

When the steam is admitted it condenses on the sphere A till the temperature reaches 100°, and the water formed by the condensation is collected in a thin catch-pan G, attached to the bottom of the sphere, and its weight is determined by putting weights on the opposite pan of the balance till equilibrium is again secured. A light metal shield H, with a hole through which the suspending wire C passes, serves to

protect the sphere from any drops of water produced by the steam condensing on the top of the vessel B. In order to prevent the condensation of steam on the wire C, where it passes through the hole in B, a spiral of fine platinum wire, I, is placed round the wire, but not touching it, and this spiral is heated by passing a current of electricity. In this way the portion of the wire passing through the hole is heated above 100°, so that no steam condenses on it.

In the best form of the steam calorimeter there is a sphere, &c., suspended from each of the arms of the balance, so that they are alongside each other in the vessel B. An experiment is then made, in which both the copper spheres are exhausted, and if they have exactly the same "water-value," the balance will remain in equilibrium after the admission of the steam. If the balance is deflected, weights are added till it comes back to equilibrium, and from the value of these added weights the difference in the water-values of the spheres can be calculated. One of the spheres is then filled with the gas to be experimented on under a pressure of about 40 atmospheres, and from the increase in the weight of the sphere the mass of the gas is obtained. The sphere is then placed in the calorimeter, the sphere attached to the other arm being still exhausted, and the steam is admitted. The sphere containing the gas now condenses more steam than the empty one, since it requires some heat to raise the temperature of the enclosed gas. The weight w, which has to be added to the other side to produce equilibrium, is then equal to the weight of the water produced by the condensation of the steam, and the latent heat given out by this steam has heated the gas in the sphere from a temperature t, say, to 100°. Hence if M is the mass of the gas, and L the latent heat of steam, the specific heat (s) of the gas is given by

$$s = \frac{wL}{M(100 - t)} \quad . \quad . \quad . \quad . \quad (68)$$

The thermal value of the copper containing-sphere does not come in since this is compensated by the empty sphere attached to the other arm of the balance.

73. Freezing- and Boiling-point of Solutions.—It has already been mentioned, when describing the determination of the lower fixed point of a thermometer, that the freezing-point of water is lowered when a salt is dissolved in water.

Careful examination of the depression of the freezing-point of different solvents produced by dissolving *small* quantities of different substances has shown that if the quantity of each substance dissolved is proportional to its molecular weight, so that in every case unit volume

of the solution contains an equal number of molecules of the dissolved substance, then the depressions in the case of any given solvent approximate to one of two constant values. In the case of solutions in water the lower of these values is due to such substances as alcohol, cane sugar, and acetic acid, while the higher value, which is approximately twice the lower, is due to such substances as sodium chloride, sulphuric acid, and hydrochloric acid.

Depression of the freezing-point.

To account for these results Van 't Hoff has advanced the theory that the depression of the freezing-point is proportional to the *number* of molecules of the dissolved substance in unit volume, and independent of the *nature* of the molecules. The lower value of the depression mentioned above may then be due to the fact that the molecules have formed into aggregates of two ordinary molecules, so that in the solution the molecular weight is doubled; or the higher value may be due to the splitting up or dissociation of the corresponding molecules when in solution.

A dissolved substance causes an elevation of the boiling-point of a solvent, and somewhat similar results have been obtained to those mentioned above in the case of the freezing-point. Although the temperature of the boiling solution is higher than that of the pure solvent, yet the temperature of the vapour over the boiling solution is the same as that over the pure solvent boiling at the same temperature. This is the reason why when determining the upper fixed point the thermometer is placed in the steam and not in the boiling water, for the presence of a very small trace of impurity will appreciably raise the boiling-point of the water.

Temperature of steam over boiling solution.

The reason the steam, say, at a little distance above a boiling solution of common salt, is at 100° C. although the temperature of the solution may be several degrees higher is as follows: The vapour which leaves the solution is the vapour of *pure* water, and at the point where it leaves the boiling solution it has a temperature equal to that of the solution. As the vapour rises it loses heat to surrounding objects, and hence its temperature falls, and as the specific heat of steam is only about ·5, a gram of steam does not require to lose more than half a calorie to allow of its temperature falling one degree. This goes on till the temperature of the steam reaches 100°, when any further loss of heat causes some of the steam to condense, but since a gram of steam in condensing to water at 100° evolves 536 calories, only a very small quantity condensing will be sufficient to allow for the lost heat, and the temperature of the steam will remain constant till it has all been condensed into water. If a thermometer is dipped into the *solution* and then suspended in the steam it will be found that it will indicate a temperature higher than 100°. The reason of this is that

the steam will condense on the solution which is clinging to the bulb and will liberate its latent heat till the temperature of the thermometer is such that it corresponds to the boiling-point of the strength of solution which is on the bulb. As, owing to the condensation of pure water on the bulb and drops of solution falling off, the strength of the solution on the bulb decreases, the temperature of the thermometer will fall, and finally when all the solution has been cleared off the temperature will become steady at 100°.

74. Freezing-point Curves for Solutions and Alloys.— Owing to the important bearing it has on the discussion of the properties of alloys, it will be desirable to examine in rather more detail what happens when a solution of common salt in water is cooled down to the freezing-

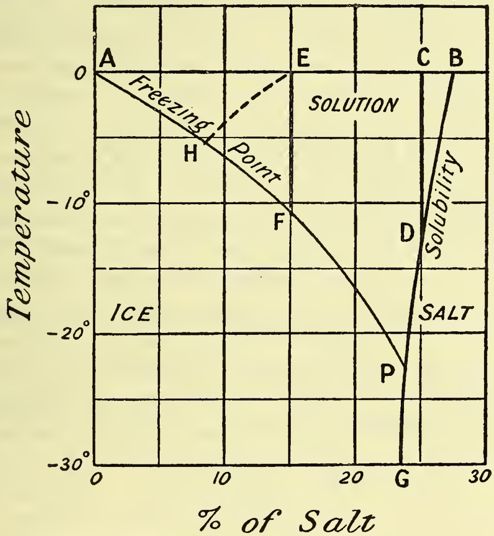


Fig. 110.

point. The amount of salt which a given volume of water can dissolve varies with the temperature, and the curve PB (Fig. 110) shows the percentage of salt in a solution which contains as much salt as can be dissolved at the different temperatures. This curve is called the solubility curve, and it will be observed that the solubility increases with rise of temperature. The curve AP gives the freezing-point of solutions of different strengths, that is, the temperature at which the solutions begin to give a solid deposit when cooled. Let us suppose that we start with a solution which contains less salt than 23·8 parts by weight of salt to 76·2 parts by weight of water, that is one corresponding to a point such as E, which is to the left of the point P where the two curves cut.

If this solution is cooled the change is represented by the vertical line EF, and down to the temperature corresponding to the point F the strength of the solution remains unaltered. If, after F is reached, the solution is further cooled, solid commences to separate, and since the solid formed is *pure* ice, the remaining solution gets more concentrated, and if the cooling is continued the change in the solution left is represented by the line FP.

Next let us start with a solution in the condition represented by the point C, that is with a solution stronger than that corresponding to the point P. As the temperature falls the change is at first represented by the vertical line CD, the strength of the solution remaining constant. If, after D is reached, the solution is cooled further it contains more salt than required to produce saturation, and hence some of the *salt* is deposited in the solid condition, and the remaining solution becomes weaker. As cooling is continued the change is represented by the curve DP. Hence, whether we start with a strong or a weak solution when the temperature has been reduced to $-22°$ C., the concentration of the solution which is still liquid is always 23·8 parts of salt to 76·2 parts of water, that is, corresponds to the point P. If the solution is further cooled, both salt and ice are deposited together in the constant proportion of 23·8 of salt to 76·2 of water, the change being represented by the vertical line PG. This constant relation between the weights of salt and ice when both are deposited together is called the *cryohydrate* of common salt.

The action of the ordinary freezing mixture, consisting of pounded ice and salt, follows from the above considerations. Suppose we start with the ice and salt at 0°. The ice has always a little water attached, and this water will dissolve some of the salt to form some solution, the state of which must be represented by some point of the diagram above the lines APB. Suppose it is a point such as E. Now the solution at E cannot be in equilibrium with ice, for it is only for points in the curve AP that such a state of equilibrium exists, and hence some of the ice will be dissolved. Owing to its latent heat the liquefaction of the ice will cause a fall of temperature. If there were no additional salt present the state of the solution would in this way change along some such line as EH. Owing, however, to the excess salt present, and the fact that solution and salt are only in equilibrium along the curve PB, salt dissolves in the solution produced by the liquefaction of the ice. This double effect, viz. liquefaction of the ice and solution of the salt, accompanied by fall of temperature, goes on till a temperature of $-22°$ C. is reached, that is till the solution is in the state represented by the point P. Since at P we can have ice, salt, and solution in equilibrium, no further liquefaction of the ice will take place, and hence the temperature will not fall any lower.

In Fig. 111 the curve ABC gives the solidifying points for alloys of silver and copper containing different proportions of the metals. The point A gives the solidifying point for pure silver, and C that for pure copper. The addition of a little copper to silver lowers the solidifying point, while the addition of a little silver to copper also lowers the solidifying point. The alloy containing 28 parts of copper to 72 parts of silver has the lowest solidifying point of all the alloys, as shown by the point B. This particular alloy is called the *eutectic*, and it plays the same part as does the cryohydrate in the case of the ice-salt mixtures considered above. If an alloy of the composition shown at P is allowed to cool slowly, when the point E is reached solidification will *start*, but the composition of the

The eutectic alloy.

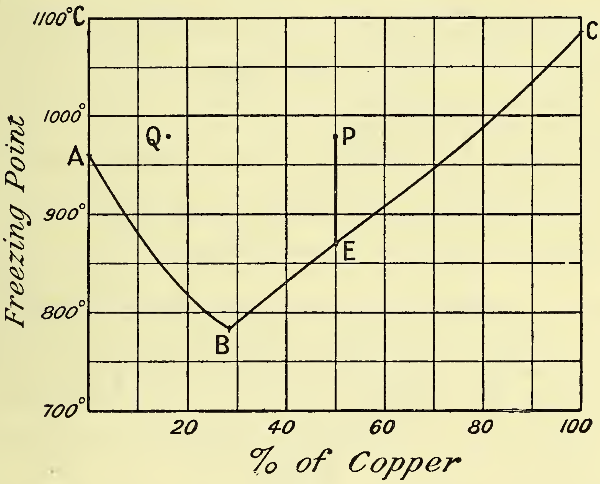


FIG. 111.

solid will not correspond to that of the alloy, but will be pure copper, and this separation of copper will continue till the point B is reached, after which the eutectic will become solid, the composition remaining constant. Thus the solid mass of alloy will consist of a mass of crystals of pure copper connected together in a matrix of the eutectic alloy. If we had started with a mixture represented by the point Q, we should in the same way, when solidification is complete, obtain an alloy which consists of crystals of pure silver embedded in a matrix of the eutectic.

The temperature at which one of the constituents of an alloy starts to solidify and that at which the eutectic solidifies can be determined by allowing the molten alloy to cool slowly, and noting the rate at which the temperature falls. Thus the curve in Fig. 112 gives the temperature

at different times of an alloy containing equal proportions of silver and copper, the temperature starting at that corresponding to the point P in Fig. 111. At first the rate of cooling is quite regular, but when the temperature falls to 870°, the rate of cooling is very small, the cooling curve becoming almost horizontal. This decrease in the rate of cooling is due to the liberation of the latent heat of the pure copper which commences to separate at this temperature.

Cooling curves.

At a temperature of 780° the rate of cooling again becomes very much reduced, this effect being due to the remainder of the alloy solidifying, *i.e.* to the solidification of the eutectic. Between the temperatures corresponding to the points E and B (Fig. 112) the alloy is partly

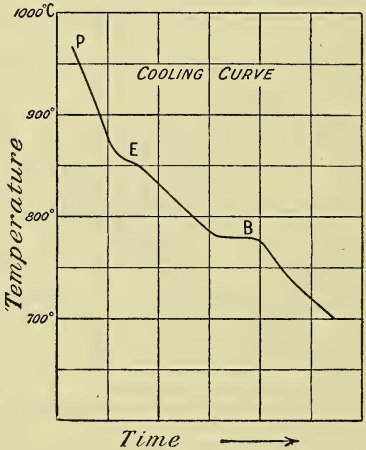


Fig. 112.

solid and partly liquid, and the points E and B indicate the solidification of pure copper and the eutectic respectively.

If the cooling curve of a sample of iron is taken, starting with the iron in the *solid* condition at a temperature of about 1000° C., the curve shows two breaks, which occur at about 860° and 750°. Since the iron is solid at the start, these pauses in the rate of cooling which indicate an evolution of heat cannot be due to a latent heat effect. They appear to be caused by molecular changes which occur in the iron at these temperatures. This explanation is supported by the fact that it is only below a temperature of 750° that iron is magnetic.

If a molten mixture of two metals is very slowly cooled, then the solid which is first separated out, say copper in the example considered on page 169, has time to form comparatively large crystals, so that the solid alloy finally produced consists of these large crystals embedded in

the eutectic. If, however, the cooling is very rapid, only very minute crystals can be produced, so that the alloy has a much finer grain. Now it is found that when an alloy is strained beyond its elastic limit it begins to yield by slip occurring along the surfaces of the crystals, and hence it will be perceived how the mechanical properties of an alloy depend not only on its composition but also on the size of the crystals contained, that is, on the manner in which it has been cooled during the process of solidification.

Crystalline constitution of solid alloys.

The copper-silver alloys which we have considered above constitute a comparatively simple case in that there is only one eutectic. In other cases, such as in gold-aluminium alloys, more than one eutectic occurs; while when we have to deal with alloys containing more than two constituents, as is the case with what are called alloy-steels, the phenomena are even more complex still. Such a steel, however, consists of an aggregate of crystals of different compositions, some of them composed of the pure constituents, and some apparently of definite chemical compounds of the constituents, and the relative quantities and sizes of the different crystals depends on the temperature to which the steel has been heated and the way it has been cooled. Thus the mechanical properties of a steel depend very largely on the heat treatment to which it has been subjected.

75. Heat Evolved during Chemical Changes.—Every chemical reaction is characterised by the evolution or absorption of a certain definite quantity of heat, so that, keeping all the external conditions the same, if the reaction takes place in the opposite sense, then the thermal phenomena simply change sign, the quantity of heat involved being the same as before. The quantity of heat involved in any given reaction depends, however, in a marked manner on the physical conditions under which the reaction takes place. Thus if 2 grams of hydrogen and 16 grams of oxygen, both in the gaseous condition, at standard pressure and temperature combine together to form water at 0°, the heat evolved by the reaction is 68834 calories. If the result of the reaction is to form steam at 100°, the heat evolved is only 57386 calories, the difference representing the heat given out by 18 grams of steam at 100° in condensing to water at 0° (*i.e.* $536 \times 18 + 100 \times 18$). In the same way, if the result of the reaction is to give ice at 0°, the heat evolved is 70274 calories (*i.e.* $68834 + 80 \times 18$).

The above is an example of a simple reaction; as a more complicated reaction, we may take the solution of metallic zinc in dilute sulphuric acid. If 65 grams of zinc are dissolved in dilute sulphuric acid: (1) water is decomposed and 2 grams of hydrogen are evolved, this reaction *absorbing* 68834 calories, as in the previous example; (2) the oxygen

combines with the zinc and 83500 calories are *evolved*; (3) the oxide of zinc combines with the acid forming $ZnSO_4$, and water and 23400 calories are evolved. Hence the resultant thermal effect of the whole reaction is that 38066 calories are evolved, since

$$38066 = -68834 + 83500 + 23400.$$

If one gram of diamond is converted into carbon monoxide (CO), 2140 calories are evolved; if the CO is then converted into carbon dioxide (CO_2), 5720 calories are evolved. Hence 7860 calories have been evolved in the conversion of carbon (in the form of diamond) into carbon dioxide, the reaction having taken place in two steps. If one gram of diamond is directly converted into carbon dioxide, the heat evolved is 7860 calories, so that the same amount of heat is evolved whether the reaction takes place in one or in two steps. This is an example of the law that when a system of bodies passes from one state to another, the quantity of heat evolved is independent of the intermediate states through which the bodies pass.

The most important chemical changes, as far as the heat developed is concerned, are those which accompany combustion, that is the combination of the substances concerned with the oxygen of the air, for it is the heat liberated during these changes which is used in all steam and gas engines. The fuels used consist mainly either of carbon or combinations of carbon with hydrogen, and during combustion the chief chemical changes involved are the oxidation of the carbon to carbon dioxide and of the hydrogen to water. The heat which is evolved when unit mass of the fuel is burnt in this way is called the *calorific value* of the fuel. If the water produced by the combustion of the hydrogen is condensed to liquid water, so that its latent heat is liberated, the heat developed by the combustion of unit mass is called the higher calorific value. In practice, however, it is generally impossible to utilise the heat liberated by the condensation of the water vapour, and hence the heat developed when unit mass of the fuel is burnt and the water vapour is not condensed is of more practical importance. This is called the lower calorific value of the fuel.

Calorific value of fuels.

If, owing to insufficient supply of air or other cause, part of the carbon in place of being converted into carbon dioxide (CO_2) is only converted into carbon monoxide (CO), then less heat will be evolved in the combustion, so that part of the heat value of the fuel is wasted. On the other hand, if the supply of air to the furnace of a boiler is too free, the supply of oxygen will be greater than that required for the combustion of the fuel, and the superfluous air will cause a waste of heat, since this air will be heated by its passage through the furnace and will carry away up the chimney a large proportion of the heat thus acquired.

CHAPTER IV

CONVECTION AND CONDUCTION

76. Convection.—When a vessel containing liquid is heated over a flame, the temperature of the portion of the liquid near the bottom of the vessel rises, and, as in general a liquid expands when the temperature rises, the density of this heated portion of the liquid becomes less than that of the superincumbent layers of liquid. A state of instability is thus set up, and as a result an upward current of the hot liquid rises from that portion of the bottom of the vessel which is over the flame, a downward current of cold liquid being also produced. Hence the heat which is communicated to the liquid in immediate contact with the heated surface of the vessel becomes distributed throughout the mass of liquid by currents set up within the liquid. These currents are called *convection currents*, and the process by which the heat is communicated is called *convection.*

Convection is made use of in the ordinary arrangement for providing a supply of hot water throughout a house, and for heating by means of hot-water pipes. In the usual arrangement the top of the boiler A (Fig. 113) is connected by a pipe BCD with a small hot-water tank E placed near the top of the house, while a return pipe F joins the bottom of the tank to the bottom of the boiler. When the water in the boiler is heated it becomes less dense and convection is produced, an upward current of hot water flowing through the pipe BCD, and a downward current of colder and hence denser water in the pipe F. Thus a room may be heated by the pipe C, or hot water drawn off by a tap as at I. To replace the water thus drawn off, the hot-water cistern E is connected with a supply cistern H by means of a pipe G. An upright tube J allows for the escape of any steam produced should the water reach the boiling-point.

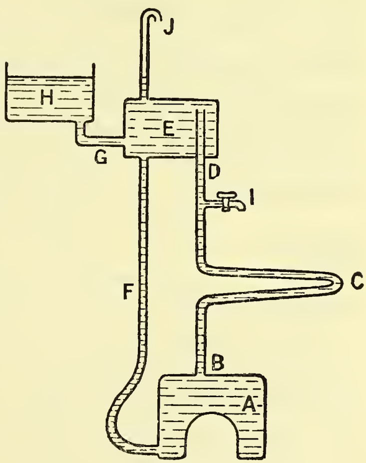


FIG. 113.

Convection is also the cause of the production of winds. Thus if the air in any given region becomes heated it expands and an upward convection current is produced, while a flow of air takes place along the surface of the earth towards the region considered. Thus a wind will be produced in the upper parts of the atmosphere which sets *away* from the heated region, and also a wind near the surface of the earth which sets towards the heated region. This effect is very clearly observable in the regular winds which are found to blow throughout the year in a definite direction over certain regions of the ocean, and are known as the trade winds. These are due to the air in the neighbourhood of the equator being heated and rising, its place being supplied by an inflow of colder air from the north and the south. If the earth were not in rotation the winds thus produced would be a north wind in the northern hemisphere and a south wind in the southern hemisphere. Owing, however, to the rotation of the earth the direction of these winds are modified. Thus consider a mass of air to the north of the equator which is travelling towards the equator. As it passes south it travels over regions of the surface of the earth for which the speed of movement due to the rotation of the earth is gradually increasing, for the linear speed of a point on the surface of the earth depends on its distance from the *axis* of rotation of the earth, and this distance is zero at the poles and a maximum at the equator. The result is that the air as it moves south lags behind the eastward motion of the surface of the earth, and hence the resulting motion of the air with reference to the surface of the earth is from the north-east to the south-west. In the southern hemisphere in the same way the trade winds blow from the south-east to the north-west.

Trade winds.

The land and sea breezes observed at some places near the coast are due to convection. During the day the land becomes more heated than the sea, and hence an upward current is produced over the land and a cool breeze blows in from the sea to take the place of the hot air which has arisen. During the night the land cools more rapidly than the sea, and hence a similar breeze is produced which blows from the land towards the sea.

Land and sea breezes.

Variable winds, gales, and cyclones are probably produced by strong convection currents set up by a more or less irregular distribution of temperature which often seems to originate in the upper portions of the atmosphere. Whenever a very low region of pressure is produced due to an upward convection current, the air rushes in from the surrounding regions, but owing to the effect produced by the rotation of the earth considered above, the path of the air is not a straight line directed towards the region of low pressure, but is a spiral forming what is called a cyclone. Thus in the northern hemisphere the wind travels towards

a region of low pressure along a spiral path, the direction of rotation being anti-clockwise. If such a wind is violent it is called a cyclone.

Cyclone.

The connection between the direction of the wind and that of the earth's surface is shown in Fig. 114, where P represents a region of *low* barometric pressure in the northern hemisphere, and Q a corresponding region in the southern hemisphere. If the pressure at a given spot is higher than that of the surroundings, there will be an outward flowing wind which again will follow a spiral path, but the direction of rotation will be reversed, that is, it will be clockwise in the northern hemisphere. Such a wind which is associated with a region of *high* pressure is called an *anti-cyclone.*

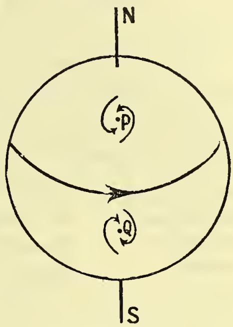


FIG. 114.

77. Conduction of Heat.—In convection the heat is communicated from one part of a body to another by the actual transfer of heated particles of the substance. Heat may, however, be communicated from one part of a body to another without any transference of any portion of matter from the hot to the cold part. Thus if one end of a rod of copper is heated it will soon be found that the far end becomes hot. In this case the heat seems to be communicated by the warmer molecules heating the neighbouring colder molecules, and so on, and the heat is said to be communicated by *conduction.*

If a slab of a given material of thickness d, with parallel faces each of area A, has its opposite faces kept at the temperatures t_1 and t_2 respectively, the quantity of heat, Q, which will pass by conduction through the slab from one face to the other in a time T is directly proportional to the area of the face, the difference in temperature and the time, and inversely proportional to the thickness. Hence

$$Q = k\frac{A(t_2 - t_1)T}{d} \quad . \quad . \quad . \quad . \quad (69)$$

where k is a constant which depends on the nature of the material, and is called *the thermal conductivity* of the material. The conductivity k is the quantity of heat which will pass in unit time between the opposite faces of a unit cube of the material, the temperature of the faces differing by one degree, for if we make T, A, d, and $(t_2 - t_1)$, each unity in the above expression, we get $Q = k$. The difference of temperature $(t_2 - t_1)$ between the opposite faces, divided by the thickness d, is called the *temperature gradient*, so that

Thermal conductivity.

the quantity of heat which passes in unit time is equal to the product of the conductivity into the area and into the temperature gradient in the direction in which the heat is flowing.

Measurement of the conductivity of copper.

An arrangement by which the conductivity of a fairly good conductor, such as copper, can be determined is shown in Fig. 115. A short and stout copper rod AB has a copper box C soldered on one end, and steam is passed into this box through the tube D. A second copper box is soldered to the other end of the rod, and is fitted with an annular partition K, so that water which flows in through G and out through I has to flow in the manner indicated by the arrows. Both boxes are covered with thick felt to prevent as much as possible loss of heat from the outside

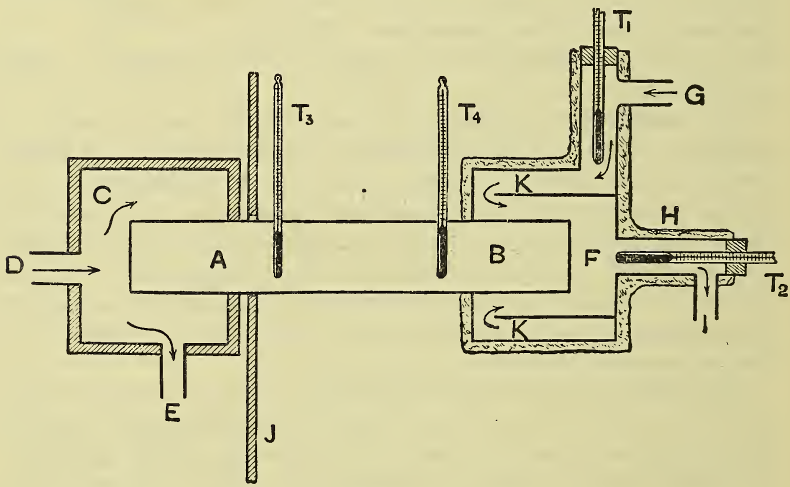


Fig. 115.

surface, while the rod itself is highly packed round with cotton-wool. Two thermometers, T_3 and T_4, are placed in holes in the rod, while two other thermometers, T_1 and T_2, serve to indicate the rise in temperature in the water which flows through the box F. After steam has been passed into C, and water through F, for some time the readings of the thermometers will become steady, and when this is the case the water which passes in a noted time is collected and weighed. If w is the weight of water which passes *per second*, and t_1 and t_2 are the temperatures indicated by the thermometers T_1 and T_2 respectively, the heat which has been conducted along the rod in a second is $w(t_2-t_1)$. If t_3 and t_4 are the readings of the thermometers T_3 and T_4, d the length of bar between the holes in which these thermometers are placed, and A

is the cross-section of the rod, we have that the quantity of heat which passes along the rod in a second is $\frac{kA(t_3-t_4)}{d}$. Now if there is no loss of heat from the surface of the rod, the heat which passes through any cross-section is the same and is equal to the heat communicated to the water. There must be a certain loss of heat from the surface between the boxes, but if the rod is *short and stout*, the proportion of the heat lost in this way to the heat communicated to the water is small, and may be neglected. Hence

$$k=\frac{w(t_2-t_1)d}{A(t_3-t_4)} \quad . \quad . \quad . \quad . \quad (70)$$

The above method is not applicable in the case of a bad conductor of heat, since with such a body the loss of heat from the surface would bear a large proportion to the quantity of heat conducted along the rod.

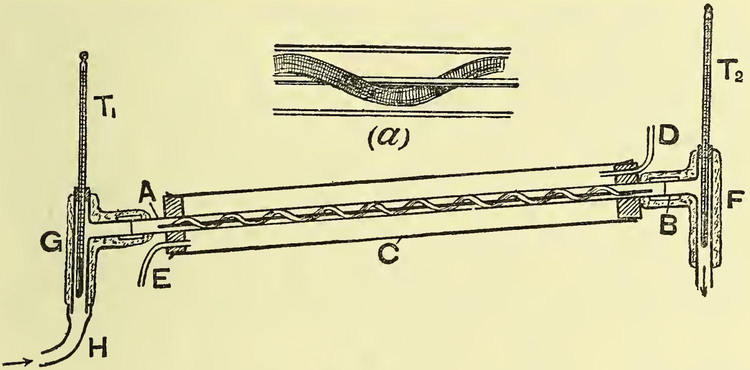


FIG. 116.

Hence another disposition has to be adopted, and an arrangement suitable for measuring the thermal conductivity of glass is shown in Fig. 116. A glass tube, AB, is enclosed for the greater part of its length in a larger tube forming a steam jacket, the steam passing in at D and out at E. Two T-pieces are connected to the ends of the tube AB, and a steady stream of cold water enters at H and escapes from F, the rise in temperature of the water being measured by the thermometers T_1 and T_2. In order to ensure that the water as it passes down the tube shall be well mixed, a piece of rubber tube, coiled round a thin rod to form a spiral, is placed in AB.

Measurement of the conductivity of glass.

The temperature, t'', of the outside surface of the glass may be taken as that of the steam, while the temperature, t', of the inside surface may be taken as half-way between the readings given by the thermometers T_1 and T_2. Let l be the length of the tube exposed to the steam, and

r_1, r_2 the internal and external radii of the tube, so that the thickness of glass through which the heat has to pass is $r_2 - r_1$. The area of the plate through which the heat passes may, if the tube is thin-walled compared to its diameter, be taken as the area, A, of a cylinder of length l and radius $(r_1 + r_2)/2$. Hence if, when the steady state is reached, w grams of water pass per second, and the readings of the thermometers T_1 and T_2 are t_1 and t_2 respectively, we have that the quantity of heat which reaches the water per second is $w(t_2 - t_1)$, and the quantity of heat which passes through the glass is

$$\frac{kA(t'' - t')}{r_2 - r_1}$$

Hence $$k = \frac{w(t_2 - t_1)(r_2 - r_1)}{A(t'' - t')} \quad . \quad . \quad . \quad (71)$$

The measurement of the heat conductivity of liquids is a difficult operation owing to the ease with which *convection currents* are set up. To reduce the formation of such currents to a minimum the liquid is heated from the top, the principle of the method being similar to that described for copper, only the cylinder of liquid is taken very wide and very short, for liquids, with the exception of mercury, are very bad conductors of heat. The small conductivity of water is of great importance in nature, since as soon as convection currents stop in a pond or lake, that is when the temperature of all the water has fallen to 4° C., all further loss has to take place by conduction through the upper layers of water and any ice which has formed on the surface. The conductivity of gases is even smaller than that of liquids, but we have seldom to do with pure conduction, since convection currents are set up so very easily in a gas. In the case of gases also it is difficult to separate the effect due to conduction from that due to radiation (§ 125). The conductivity of gases is independent of the pressure except at *very low* pressures.

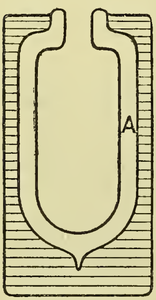


FIG. 117.

The "warmth" of woollen clothing is almost entirely due to the poor conducting-power of air, for owing to the air being entangled with the wool fibres convection currents are not easily set up, so that the heat has to pass from the body to the outside air by conduction through the wool fibres and through the enclosed air.

In the absence of all matter *conduction* cannot take place, and hence a very complete vacuum transmits no heat by conduction, all heat which passes being due to radiation. The vessels designed by Dewar to contain liquefied gases and the ordinary thermos flask depend for their

action on this fact that a vacuum cannot conduct heat. The liquid which is to be kept hot or cold, as the case may be, is enclosed in a double-walled glass vessel A (Fig. 117), the space between the walls being very carefully exhausted. Since, as we shall see later, a bright metallic surface radiates badly, the inside surface is silvered, as by this means the loss of heat by radiation is decreased. The felt packing round the outside of the flask is of little service in preventing the transfer of heat; its function is chiefly a mechanical one of protecting the somewhat fragile glass vessel from injury.

Vacuum flask.

In the case of the Davy safety-lamp for use in coal mines, the fact that metals are good conductors of heat is utilised. In some mines so much inflammable gas is evolved from the coal that the presence of a naked flame would cause an explosion. Since, however, an inflammable mixture of gases has to be raised to a fairly high temperature before combustion occurs, if by any means the gases can be kept sufficiently cold no flame can be propagated. In the Davy lamp the air supply to the oil flame passes through some fine wire gauze, and although the inflammable gas may combine inside the lamp, yet the flame cannot strike back through the gauze, since the gas in the interstices of the gauze is too cold, any heat locally produced being immediately conducted away by the metal.

Davy safety-lamp.

In the case of the fire-box of a steam boiler the heat of the flame has to pass through the metal plate into the water by conduction. Hence the rate at which heat passes depends on the conductivity of the metal. In practice, however, the flow of heat is governed more by the presence of a thin coating of scale and soot on the fire side, and of grease and "fur" on the water side than by the thickness of the metal. The reason is that thermal conductivity of such coatings are enormously greater than that of iron or steel. In considering the passage of heat from the hot *gas* in the furnace to the *water* in the boiler, it is important to remember that a great deal of the resistance to heat transfer is in the passage from the gas to the metal on one side and from the metal to the water on the other. After, say, the gas in immediate contact with the metal has parted with some of its heat, unless this layer is immediately swept away the heat from the rest of the gas will have to pass by conduction through this layer of badly conducting gas. A similar effect takes place on the water side. The result is that the rate at which heat travels from the furnace to the water depends more on the scouring action by which the film of gas or water in contact with the metal is replaced than on the conductivity of the metal.

CHAPTER V

CONNECTION BETWEEN WORK AND HEAT

78. Mechanical Equivalent of Heat. First Law of Thermodynamics.—It is a familiar fact that heat can be produced by rubbing one body against another, that is, by doing work against friction heat is produced. Further, it is well known that by means of a steam-engine work can be performed by the heat generated when coal is burnt. The above knowledge of a qualitative relation between work and heat was extended by Joule, who first showed by direct experiment that there exists a definite quantitative relation between the heat and the work when one of these is converted into the other. He showed that whenever mechanical energy is converted into heat, or heat into mechanical energy, the ratio of the quantity of mechanical energy to the quantity of heat involved is constant. The above result is known as the first law of thermodynamics, and may be expressed in symbols as follows: If W units of work are converted into heat, or *vice versâ*, the quantity of heat produced, H, is given by

First law of thermodynamics. Mechanical equivalent of heat.

$$W = JH \qquad . \quad . \quad . \quad . \quad (72)$$

where J is a constant independent of the manner or direction in which the transformation takes place. The quantity J represents the amount of work, that is *energy*, which must be converted into heat to produce one heat unit, and is called the *mechanical equivalent of heat.*

The numerical value of J obviously depends on the values of the units of work and heat employed. Thus in the C.G.S. system $4 \cdot 189 \times 10^7$ ergs of work have to be expended to produce one calorie. In the British system 778 foot-pounds are equal to one British thermal unit (B.Th.U.).

Since work and heat are intimately related, as indicated by the first law of thermodynamics, it is natural to speculate as to what is the nature of heat. We have, in § 17, considered a case where a similar relation exists, namely, that between the kinetic energy of a body and the work it can do when losing its motion. Hence we are led to consider whether the energy possessed by a stationary hot body, in virtue of its heat

may not be of the nature of kinetic energy. Since the hot body is not in visible motion, the motion involved must be that of the molecules, or even atoms, which constitute the body. There are many other phenomena which support this view, one of which may be mentioned. When a body, such as a spring, is in rapid vibratory motion in the air, it produces waves in the air which constitute sound. In the same way when a body is heated it sends out waves into the surrounding space, which waves, if they satisfy certain conditions which we shall consider later, affect the human eye and produce the sensation of light. Hence just as the moving spring produces sound-waves in the air, so do the moving molecules of a hot body produce light-waves.

When we are dealing with mechanical energy and also heat, it is often convenient to replace the ordinary heat units (calories or B.Th.Us.) by energy units (ergs or foot-pounds), since in this way the use of the factor J is avoided. Thus in the following pages, where we are dealing with transformation of mechanical energy into heat, and *vice versâ*, we shall in general suppose the heat to be measured in mechanical units.

Suppose that a quantity of heat H (measured in mechanical units) is communicated to a body, but during the process of heat supply the body does an amount of external work W. Then on account of the heat supplied the energy of the body has increased by an amount H. Since, however, the body has performed the external work, it has lost energy to the extent of W. Hence the net *gain* of energy is $H-W$, and if E_0 is the energy possessed by the body before the change and E that possessed after, we have

$$H-W=E-E_0$$

The quantity E is called the *intrinsic energy* of the body, and writing the above equation

$$H=W+E-E_0 \quad . \quad . \quad . \quad . \quad (73)$$

Intrinsic energy of a body.

we see that the heat communicated to a body in any change is equal to the external work performed by the body during the change together with the change in the intrinsic energy. This result is really a combination of the first law of thermodynamics with the principle of the conservation of energy.

We cannot determine the actual value of E for a body, since we are unable to extract all its energy and measure the amount, but as we are concerned only with the *change* in E, no inconvenience is caused on this account. It is usual to assume that the intrinsic energy of a body is zero at 0° C., and calculate all changes of intrinsic energy from this zero.

Joule determined the value of the mechanical equivalent by measuring the heat produced by friction in a calorimeter when the liquid is rapidly stirred, the work of stirring being performed by falling weights. His method was much improved by Rowland, and hence the arrangement adopted by this observer will be described. The work is converted into heat by friction produced in the water contained in a calorimeter by a paddle-wheel BB, shown in plan in Fig. 118, the arms of which passed through slots in the fixed vanes C, the object of the vanes being to prevent the water being set in rotation by the motion of the paddle. The paddle-wheel was driven by a belt FE passing round a pulley A, while the calorimeter was supported by a fine wire so that it was capable of rotating about an axis parallel to that about which the

Determination of the value of the mechanical equivalent.

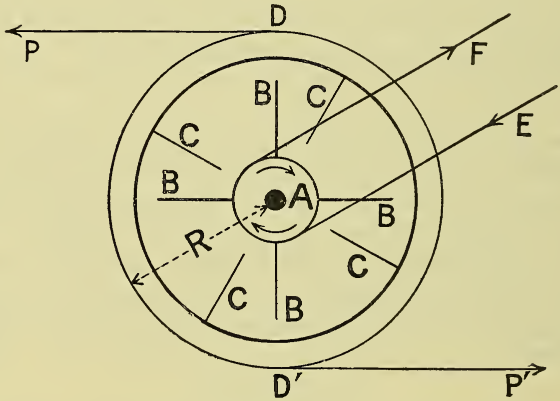


FIG. 118.

paddle rotated. Owing to the motion of the paddle the calorimeter tends to turn in the same direction, and to prevent this two strings DP, D′P′ were attached to the circumference of a disc which was fixed to the calorimeter. These strings passed over light pulleys, and equal weights were suspended at the ends, the weights being just sufficient to prevent the calorimeter turning when the paddle was rotated at a given speed. If w grams is the weight of each of these suspended weights, the pull exerted along DP and along D′P′ is wg dynes, and if the radius of the disc DD′ is R, the couple exerted on the disc is $2Rwg$. Since the calorimeter is in equilibrium and action and reaction are equal and opposite, it follows that the turning couple exerted by the paddle on the calorimeter is $2Rwg$, and this must be the retarding couple exerted by the calorimeter on the paddle. Now the work done against this couple

during one complete rotation of the paddle is $2\pi . 2Rwg$ (§ 25). Hence if the paddle makes n revolutions, the work done by the paddle on the water contained in the calorimeter is $4\pi nRwg$. Hence if the water value (§ 66) of the calorimeter and *all* its contents is C, and the rise in temperature during the time the paddle makes n revolutions is t, we have

$$J = \frac{4\pi nRwg}{Ct} \quad . \quad . \quad . \quad . \quad (74)$$

In a later section (§ 154) a method of determining J by the heat developed in a wire when it is traversed by an electrical current will be described.

79. Work done by a Gas during Expansion.—Suppose we have m grams of a gas at an absolute temperature T_1 enclosed in a uniform cylinder of cross-section A, by a weightless piston B (Fig. 119), which can move without friction within the cylinder. Further, let there be a pressure of P dynes per square centimetre acting on the upper side of B. Since B has no weight and moves without friction, the pressure exerted by the gas on the under side of B must also be P. Now let us heat the gas to a temperature T_2. The pressure to which the gas is subjected will remain the same, and hence the gas will expand, driving the piston back to the position B′. If the distance of the piston from the bottom of the cylinder at the temperature T_1 is h_1, and that T_2 is h_2, then the distance through which the piston has been driven back is $h_2 - h_1$. Since the force acting on the upper surface of the piston is PA, the work performed as the piston is driven back is $PA(h_2 - h_1)$. But Ah_1 is the volume occupied by the gas at T_1, and Ah_2 is the volume occupied at T_2, so that $A(h_2 - h_1)$ is the increase in volume of the gas. **Thus the external work done by the gas in expanding at constant pressure is equal to the product of the pressure into the change in volume.** The same amount of work would be performed even if the piston were not present, or even if the gas were not confined in a cylinder, so that the above result is quite general. If the gas *decreases* in volume, work is done *on* the gas by the external pressure.

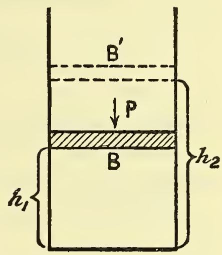


FIG. 119.

We have now to consider how much heat has to be supplied to the gas while the temperature is raised from T_1 to T_2 at a constant pressure P. If m is the mass of the gas and v is the *increase* in volume, the external work performed by the gas is Pv, while the heat which has to be supplied is $mC'_p(T_2 - T_1)$ where C'_p is the specific heat at constant

pressure (§ 66) expressed in mechanical units. Hence equation (73), p. 181, gives

$$mC'_p(T_2 - T_1) = Pv + E_2 - E_1 \quad . \quad . \quad . \quad (75)$$

where E_1 and E_2 are the intrinsic energies of the gas at T_1 and T_2 respectively.

Next suppose the same quantity of gas is heated from T_1 to T_2, the pressure starting at P, but the volume V being kept constant. If C'_v is the specific heat at constant volume expressed in mechanical units, the heat which has to be supplied is $mC'_v(T_2 - T_1)$, and since no external work is performed,

$$mC'_v(T_2 - T_1) = E'_2 - E_1 \quad . \quad . \quad . \quad (76)$$

where E'_2 is the intrinsic energy of the gas when at a temperature T_2 but at the volume V, and E_1 has the same value as before. The question now arises, is E_2 equal to E'_2. The temperature in the two cases is the same, the only difference being that for E_2 the volume is $V+v$, while for E'_2 it is V. If the molecules of the gas do not attract each other, then no work has to be done to separate them, or in other words, the intrinsic energy of a given mass of gas at a given temperature is independent of the distance between the molecules, that is, of the volume occupied by the gas. Now experiment has shown that only *very little* work has to be done to separate the molecules of most gases. Hence, as a near approximation we may put $E_2 = E'_2$, whence equations (75) and (76) give

$$mC'_p(T_2 - T_1) - P_v = mC'_v(T_2 - T_1)$$

or

$$C'_p - C'_v = \frac{P_v}{m(T_2 - T_1)} \quad . \quad . \quad . \quad (77)$$

But if C_p and C_v are the specific heats of the gases in thermal units, we have $C'_p = JC_p$ and $C'_v = JC_v$. Hence

$$J = \frac{P_v}{m(C_p - C_v)(T_2 - T_1)}$$

But since V and $V+v$ are the volumes of the gas under constant pressure at the absolute temperatures T_1 and T_2, we have (§ 63)

$$\frac{V+v}{V} = \frac{T_2}{T_1}$$

or

$$\frac{v}{V} = \frac{T_2 - T_1}{T_1}$$

Hence, substitute for v in the equation above,

$$J = \frac{P}{T_1(C_p - C_v)} \cdot \frac{V}{m}$$

But m/V is the density D of the gas at the pressure P and temperature T,

$$\therefore \qquad J=\frac{P}{T_1(C_p-C_v)D} \quad . \quad . \quad . \quad . \quad (78)$$

It will thus be seen that if we know the specific heats of a gas for which we know the density at a given pressure and temperature we can calculate the value of the mechanical equivalent. This is Mayer's method of obtaining the value of J, and it is only valid if the hypothesis we have made, namely, that $E_2 = E'_2$, is valid.

Value of J deduced from the difference in the specific heats of a gas.

For air, $C_p = 0{\cdot}238$ calories, $C_v = 0{\cdot}169$ calories, and at a pressure of one atmosphere, or 1013300 dynes per square centimetre, and a temperature of 273° absolute (0° C.), the density is 0·001293. Hence, substituting these values in the equation above, we get

$$J=\frac{1013300}{273\times{\cdot}069\times{\cdot}001293}=4{\cdot}16\times10^7$$

To investigate whether when a gas expands work is done in separating the molecules, Joule and Thomson (Lord Kelvin) allowed a gas to expand without doing external work. If in expanding work is done in separating the molecules, then, by the first law of thermodynamics, heat must be used up, *i.e.* the temperature of the gas will fall. The principle of the method employed in order to allow a gas to expand without doing external work is as follows :—

Measurement of work performed against molecular forces when a gas expands.

Let the original pressure and volume of unit mass of a gas be p and v, and the final pressure and volume p' and v'. Further, let the passage of the gas from one state to the other be made by means of the arrangement shown in Fig. 120, in which A and B are two pistons connected by a rod which passes air-tight through a partition separating the spaces C and D, and which move without friction in two cylinders, the cross-sections s and s' of these cylinders being in the ratio of v to v'. If the spaces C and D between the two pistons are vacuous, the work done by the gas on A while it moves through unit distance to the right is ps, while the work done by B in pushing the gas forward is $p's'$. Hence the difference between these two quantities of work is $ps-p's'$, or, since $s/s'=v/v'$, this difference is proportional to $pv-p'v'$. Now if Boyle's law holds for the gas $pv=p'v'$, hence on the whole no work is done on or by the double piston. Next let the space C be filled with gas at the pressure p, and the space D with gas at the pressure p', and let these two spaces be connected by a tube, E, in which is a diaphragm, F, pierced with a very small hole. The gas will gradually pass through this hole,

and, as is evident, if the double piston is moved so as to keep the pressure in C constant and equal to p, the pressure in D will also be constant and equal to p'. When each piston has passed through unit distance, a certain mass of the gas will have passed from C to D, its pressure changing in the process from p to p'. The gas escaping into D has done no work in forcing the piston B back, since the pressure of the gas acting on A will, as we have seen, exactly do the requisite work. This energy is of course supplied by the pump used to keep the pressure to the left of A constant, which process might be performed by a second piston, G, working in the cylinder and driven forward by hand. Hence we have allowed the gas to expand without doing external work, and any change of temperature it experiences must be due to the performance of internal work.

The same process would go on if the pistons were not present, for throughout the change we have supposed the pressure on the two sides

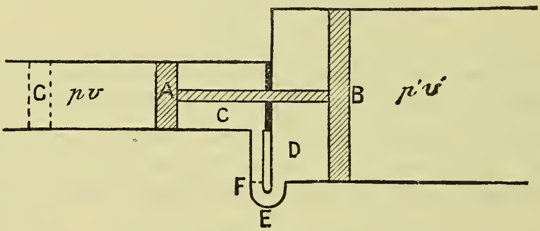


FIG. 120.

of each piston to remain the same, so that if gas is allowed to escape through a fine opening, any change in temperature produced will be due to internal action between the molecules. The temperature must not be taken immediately at the opening, for there the gas, as it rushes out, possesses considerable kinetic energy, and it is only after this kinetic energy has been lost by the friction of the gas against itself and against the walls, &c., and the heat energy originally used up in setting the gas in motion is returned to the gas in the form of heat, that no external work has been done on the expanding gas.

In their experiments, Joule and Kelvin allowed a steady stream of gas to pass through a long copper spiral immersed in a water bath kept at a uniform temperature. The gas then escaped through a porous-plug made of cotton-wool, which acted the part of the fine hole F, and also prevented the gas from leaving with any appreciable kinetic energy, since the gas rapidly loses its velocity as it passes through the interstices of the wool. The temperature of the gas before and after its passage through the plug was indicated by two delicate thermometers.

In the case of most gases a small drop of temperature was obtained, indicating that a certain amount of work has to be done in these gases to separate the molecules, so that using the notation employed earlier, E'_2 is slightly greater than E_2. In the case of hydrogen at ordinary temperatures there was a slight rise of temperature, but subsequent experiments have shown that at temperatures below about $-150°$ C. there is a cooling in the case of this gas also.

The change in temperature obtained in Joule and Thomson's experiment is quite small, the drop in the case of air being about 0·25° C. per atmosphere difference in pressure on the two sides of the plug. Small as this fall of temperature is, it has been utilised by Linde and by

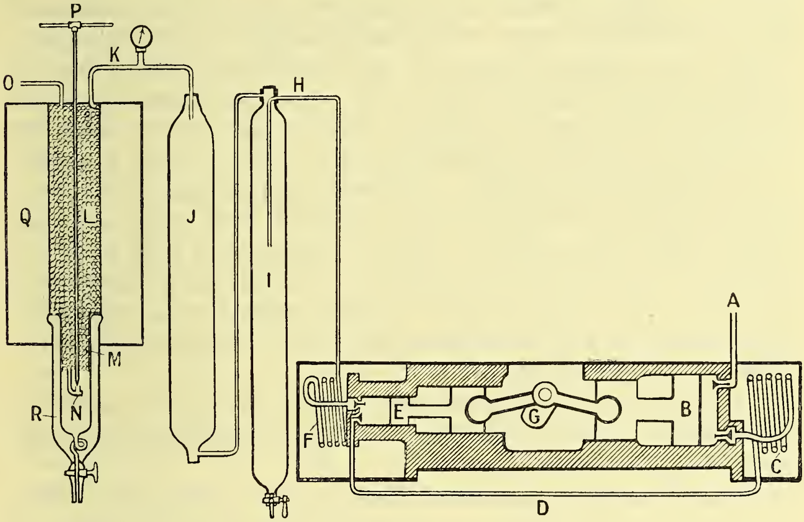


FIG. 121.

Hampson in their methods for the liquefaction by *cold* of air and other of the so-called permanent gases.

Liquefaction of air.

In Hampson's form of the liquid-air plant, the air, together with a little water to act as a lubricant, is drawn in through a tube A (Fig. 121), and compressed by the pump B to about 14 atmospheres. It then passes through a coil of copper pipes C immersed in cold water, where the heat produced by compression is removed from the gas, which then passes to the second pump E through the tube D. This pump compresses the air to about 200 atmospheres, and the highly compressed gas is first cooled in the water-cooled coil F and then passes through two strong steel reservoirs I and J. In I most of the moisture carried by the gas is deposited,

while J, being packed with sticks of caustic potash, removes the remainder of the moisture and any carbon dioxide present in the air. The air then passes through the tube K to a coil L, consisting of a very large number of turns of fine copper tube, escaping through a fine orifice at N. The size of this orifice can be adjusted by a needle valve operated by the rod P. The air having been reduced to very nearly atmospheric pressure at the valve, passes up between the spirals of copper L, and finally escapes through the tube O. When the air escapes through the fine hole at the valve N, a cooling takes place just as in the Joule-Kelvin porous-plug experiment. This cooling amounts to about 50° C. The cooled air as it flows up past the copper spirals abstracts heat from them, and when the air finally escapes at O its temperature has almost fallen to that of the compressed gas entering through K. Owing to the cooling produced in this way by the escaping gas, the compressed gas at the lower part of the spiral gets colder and colder, a further cooling always taking place when it expands at the nozzle, till finally the temperature is sufficiently low for some of the air to liquefy.

The liquid air collects in the Dewar vacuum jacketed vessel R, whence it can be drawn off by a tap as required. The upper part of the copper spiral is enclosed in a cylinder Q packed with wool, which is a very bad conductor of heat, while the lower part M of the spiral is enclosed in the vacuum vessel R. Such a machine is capable of producing from 1 to 1·5 litres of liquid air per hour, the power required to drive the pumps being about 5 HP.

When used to liquefy gases other than air, the pipes O and A are joined and an auxiliary pump forces in enough additional gas to make up the loss due to liquefaction.

Since at ordinary temperatures there is a *rise* of temperature when hydrogen is expanded through a nozzle, the above method would not work. However, if the hydrogen is cooled by passing through a spiral immersed in liquid air before it reaches the spiral L (Fig. 121), the temperature may be so reduced that on expansion a *cooling* takes place, and then the process proceeds as in the case of air.

External work performed when a liquid is vaporised.

When a liquid is converted into a vapour at constant pressure, as occurs when a liquid boils, the volume increases very greatly and a considerable amount of external work is performed, the equivalent amount of heat forming part of the latent heat of vaporisation. Since one gram of water at 100° C. increases in volume by 1649 c.c. when becoming steam at the same temperature and a pressure of one atmosphere (1013300 dynes per square centimetre), the external work performed is 1013300×1649 ergs, and this is equal to $\dfrac{1013300 \times 1649}{4{\cdot}2 \times 10^7}$ or 39·9

calories. Thus of the 536 calories required to convert a gram of water at 100° C. into steam at the same temperature, 39·9 calories are expended in performing external work and 496·1 calories are expended in doing the internal work of separating the molecules.

80. Graphical Representation of the connection between the Pressure and Volume of a Substance. Isothermal and Adiabatic Curves.—If we are given the volume occupied by unit mass of any substance, that is the specific volume, together with the pressure and temperature, then the state of that substance is defined. Thus suppose we are told that the specific volume of water stuff is unity, when the temperature is 4° and the pressure is one atmosphere, then we know that the water must be entirely in the liquid condition. If, however, at the same pressure and temperature we are told that the specific volume is 2, it necessarily follows that the water stuff exists partly as liquid and partly as gas, and it is evident that on occasion it might be useful to have some means by which, given the pressure p, the specific volume v, and the absolute temperature T, we could at once find the proportion of liquid and gas present. This information might be embodied in tables, but if they are to cover all possible combinations of p, v, and T they would have to be very extensive. Since there are the three variables p, v, and T to consider, we cannot draw a plain curve representing the variation of these three quantities, since on a plane curve we can only show the connection between two quantities. If, however, we suppose one of the three, say the temperature, to remain constant, we can draw curves which shall give for any substance the relations between p and v for any given temperature. Further, if we draw a series of such curves for different temperatures, say for every degree or ten degrees, we can by interpolation find the relation between p, v, and T. Evidently we can in this way draw three series of curves for any given substance, according to whether for each curve T, p, or v is kept constant. Which series is the most convenient depends on the problems which are being discussed. For many practical purposes the curves corresponding to constant temperature, which are called *isothermals*, are the most convenient, and hence we shall proceed to consider some of the properties of these curves.

Isothermal curves.

In the case of a perfect gas, that is one which obeys Boyle's law, so that if the temperature is constant, the product pv is constant, the isothermals have the form[1] shown in Fig. 122. In this figure the isothermals for air are given for every 100° from ·0° C. to 500° C. If the state of a gas is represented by the point A and the gas expands, the temperature being

[1] The curves, of which the equations are pv=constant, are rectangular hyperbolæ of which the axes of pressure and volume are the asymptotes.

constant, the connection between the pressure and volume is represented by the isothermal through A. If the expansion is stopped at the point B, the gas will have performed external work, and this work will be represented by the area of the figure MABN (§ 15). Now since this external work has been performed, heat must have been used up to an equal amount (first law of thermodynamics). Hence since by hypothesis the temperature of the gas has been kept constant during the change, so that this heat cannot have been derived from the heat possessed by the

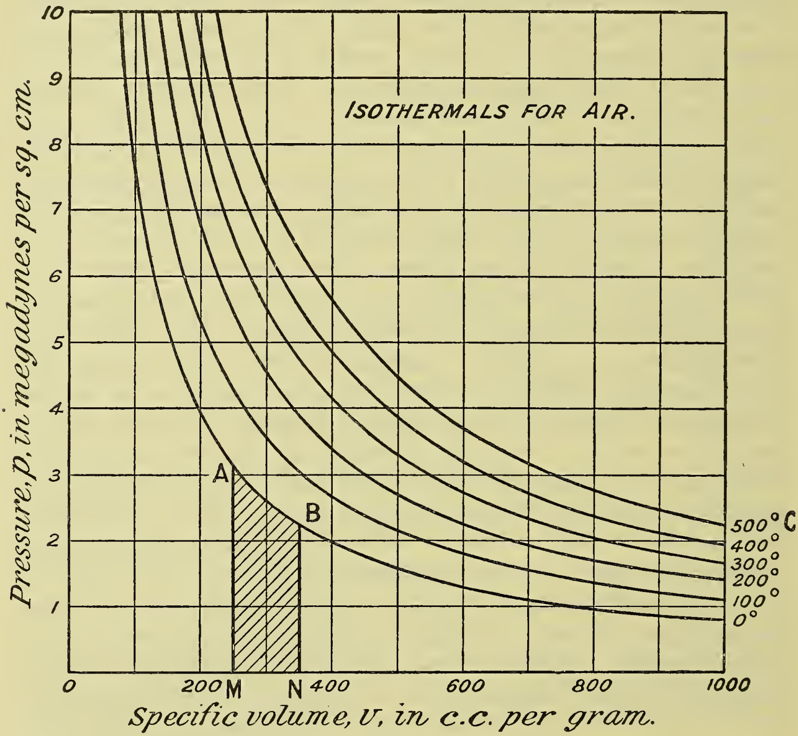


FIG. 122.

gas, heat to an amount represented in mechanical units by the area of the figure MABN must have been supplied to the gas from external sources. Thus a gas can only expand isothermally if heat is supplied to it. If the gas is compressed isothermally from B to A, then work is performed *on* the gas and heat must be abstracted from the gas in order to secure that the process shall be isothermal, the amount of this heat being represented by the area of the figure NBAM.

We have now to consider the form of the isothermals for a liquid

and its vapour, and to trace the change from the liquid state through that of a saturated vapour to that of an unsaturated vapour.

The general form of the isothermals for water and steam are shown diagrammatically in Fig. 123. Suppose we had a gram of water contained in a cylinder closed by a piston, and that the temperature, pressure, and volume of the water are represented by the point A. Further, suppose that the cylinder is kept at a constant temperature T_1, and that the load on the piston is gradually decreased, so that the pressure to which the water is subjected is reduced. As a result the volume of the water will increase, the change being represented by the line AB. At the point B where the pressure p_b is equal

Isothermals for liquid and vapour.

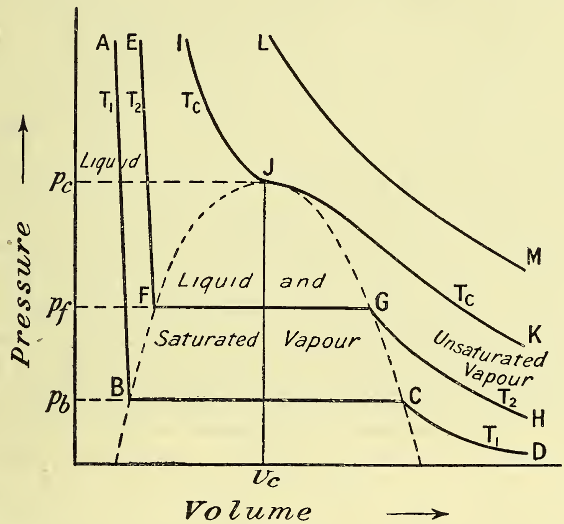


FIG. 123.

to the vapour tension of water at the temperature T_1 the liquid commences to vaporise, and if the piston is drawn back so as to increase the volume the pressure will remain constant till the whole of the liquid has been converted into steam. Thus the isothermal corresponding to the gradual change from water to steam will be represented by a horizontal line BC. For all points in BC the cylinder will be filled partly with water and partly with steam, *i.e.* saturated vapour, the proportion of water to steam decreasing as we go from B to C. At the point C the whole of the water will have become vapour, and if the pressure is further reduced the isothermal has the form CD, and corresponds to that of an unsaturated vapour. Since an unsaturated vapour which is far removed from its point of saturation behaves like a perfect gas, the

isothermal beyond D resembles in shape those shown in Fig. 122. Next let us start with the water at a higher temperature T_2. Since the temperature is higher the specific volume at any pressure will be larger than before, and the isothermal for the liquid will be EF. Further, since the temperature is higher the vapour tension p_c is also higher, and hence the liquid commences to vaporise at a point F which corresponds to a higher pressure and a larger specific volume than before. Along FG we have both liquid and vapour coexisting, while beyond G we have only vapour. Experiment shows that the density of saturated steam increases with increase of the pressure, hence since the points C and G correspond to saturated steam, the density at G is greater than that at C, that is, the specific volume at G is less than that at C. Hence, since F lies to the right of B and G to the left of C, the horizontal portion of the isothermal decreases in length as the temperature rises, till finally for a temperature T_c there is no horizontal part in the isothermal IJK. If we draw a curve through the points BFJ, &c., where the liquid commences to vaporise, and one through the points CGJ where the vaporisation is just complete, we get a curve, shown dotted in the figure, such that for all points enclosed we have both liquid and vapour coexisting. For all points to the left of this region corresponding to temperatures lower than T_c we have liquid only, and for all points to the right corresponding to temperatures lower than T_c we have vapour only, and the change from liquid to vapour is quite marked, and corresponding to the horizontal portion of the isothermals both liquid and vapour are present with a definite surface of demarcation between them, so that if we have a gram of water in a glass tube of which the volume was v_c, the pressure being p_f and the temperature T_2, the lower part of the tube would be occupied by water and the upper part by vapour, the surface of separation being clearly visible. If, however, the temperature of the tube were raised to T_c, so that the conditions are represented by the point J, it would be found that the tube was filled with a homogeneous fluid. If keeping the temperature constant we could increase the pressure, the contents of the tube would become definitely liquid, while if the pressure were decreased we should find the contents of the tube to have the properties of a gas, but the change from liquid to a gas would be a gradual one, and we could not say the change took place at any definite point. Thus at all temperatures above T_c the distinction between the liquid and the gas disappears. The

Critical point.

temperature T_c is therefore called the *critical temperature* for the given substance, and it is only at temperatures below the critical temperature that we can have the substance in the form of a liquid and vapour coexisting. Thus if we require to liquefy a gas, that is, to obtain a liquid with a definite surface of demarcation separating it from the unliquefied gas, we must work at a temperature

which is lower than the critical temperature. This condition was first discovered by Andrews, who showed that unless the temperature was below the critical temperature no increase in pressure however great was sufficient to liquefy a gas. The pressure p_c and the specific volume v_c corresponding to the point J are called the critical pressure and volume respectively.

The isothermals for temperatures well above the critical temperature have the same form throughout as those of a gas, as shown at LM (Fig. 123). The actual form of the isothermals for carbon dioxide, as

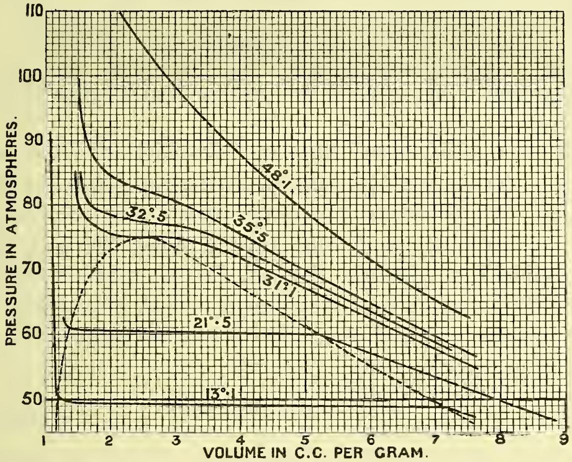


FIG. 124.

determined by Andrews, for the different temperatures written alongside are shown in Fig. 124.

When studying the isothermals we have considered the changes in pressure and volume of a substance when the temperature is kept constant, and we now have to investigate the relation between these quantities subject to a different condition which is of great practical importance to engineers, namely, when no heat is supplied to, or removed from, the substance during the change. A curve showing the relation between the pressure and specific volume when no *heat* is allowed to enter or leave the substance is called an *adiabatic* curve.

Adiabatic curves.

It follows from what has been said on page 183, that when a body expands and does external work, then, if no heat is supplied to the body, its temperature must fall, while if under the same conditions the

body is compressed, the temperature will rise. If AB (Fig. 125) represents a portion of the isothermal through a point P, then if, starting with the substance in the conditions indicated by P, we compress the substance adiabatically, we do work on it, and therefore its temperature will rise, and for a given pressure the volume will be greater than it would be if we had kept the temperature constant, *i.e.* travelled along the isothermal PA. Hence the adiabatic CD through P is more steep than the isothermal through the same point.

As we have seen, the equation to an isothermal for a perfect gas is $pv=$ constant. It can be shown that the relation between the pressure and volume during an adiabatic change is such that

$$pv^{\gamma} = \text{constant} \quad . \quad . \quad . \quad . \quad (79)$$

where γ (gamma) is the ratio of the specific heat at constant pressure to that at constant volume (§ 66).

Adiabatic elasticity of a gas.

In § 40 we have shown that the elasticity of a gas at constant temperature (isothermal elasticity) is equal to the pressure. We have now to consider what will be the elasticity under adiabatic conditions, that is, when the heat produced by the compression is retained in the gas. If the pressure to which a gas is subjected is increased by a *small* amount from P to $P+p$, while as a result the volume changes from V to $V-v$, we have, if the change is adiabatic, from equation (79),

$$PV^{\gamma}=(P+p)(V-v)^{\gamma}$$

or expanding by the binomial theorem and neglecting terms involving the square or higher powers of the small quantity v,

$$PV^{\gamma}=(P+p)(V^{\gamma}-\gamma V^{\gamma-1}v)$$
$$=PV^{\gamma}-\gamma PV^{\gamma-1}v+pV^{\gamma}$$

since the term involving the product of the small quantities p and v can be neglected. Hence

$$\gamma P=\frac{pV}{v}$$

But in § 40 it was shown that the elasticity E of a gas is equal to pV/v. Hence

$$E=\gamma P \quad . \quad . \quad . \quad . \quad . \quad (80)$$

That is, the adiabatic elasticity is equal to the pressure multiplied by the ratio of the specific heats.

As an illustration of the kind of information we may derive from the

form of the adiabatic curves for a substance, we will consider what happens when saturated steam is compressed or expanded. Suppose we have unit mass of *saturated* steam in the condition represented by the point A (Fig. 126), and that CAB represents the isothermal through A, and RAP is the adiabatic drawn through A, while the dotted curve QAS is part of the saturation curve CGJ (Fig. 123). Then if the steam is compressed *isothermally*, it will immediately commence to condense, for the change produced would be along the line AB. If, however, starting from the same point A, we compress the steam adiabatically, the change will be indicated by the portion of the adiabatic drawn through A, that is by the curve AP (Fig. 126). Thus after compression without loss of heat the condition of the steam is represented, say, by the point P, and this point indicates that the steam is now *unsaturated*, for at the pressure corresponding to P

Expansion of saturated steam.

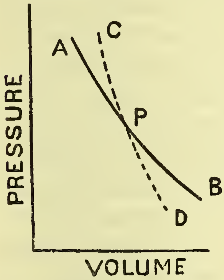


FIG. 125.

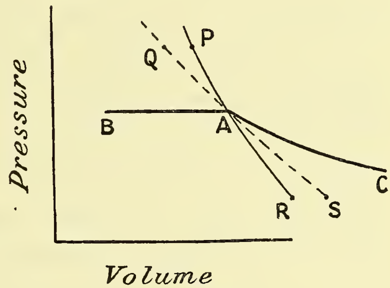


FIG. 126.

the specific volume of *saturated* steam would be that corresponding to the point Q on the saturation curve QAS. On the other hand, if we had expanded the saturated steam adiabatically to the point R, some of the vapour would condense, for R is in the region corresponding to a mixture of vapour and liquid.

It will thus appear that compressing steam without loss of heat will supply so much heat, due to the work done on the steam, that the steam becomes superheated, *i.e.* unsaturated. If, however, we abstract some of the heat produced by the work done on the steam, we could arrange that the change in pressure and volume takes place along the saturation curve AQ, *i.e.* that the steam remains *saturated* throughout. If the temperature between A and Q is one degree, then the amount of heat abstracted in this way from unit mass of saturated steam is called the specific heat of *saturated steam*, and since the heat has to be abstracted, although the temperature rises this specific heat is said to be negative. It must be carefully borne in mind that

Specific heat of saturated steam.

heat is really supplied to the steam to raise its temperature, this heat being derived from the external work performed by the agent that compresses the steam. In order, however, to just keep the steam saturated, too much work is performed, and hence the *excess* heat has to be removed. In the same way, when saturated steam expands adiabatically and the temperature falls, the steam does so much external work that unless some additional heat is supplied from outside, some of the steam must condense, the latent heat thus liberated supplying the necessary heat.

Cloud formed by adiabatic expansion of saturated air.

As a result of the above properties of steam, if saturated air in a vessel is *suddenly* expanded, so that the air has not time to take up heat from the walls, the air will become supersaturated, and if there is any dust present in the air, some of the water will be separated in the form of a cloud. In the complete absence of dust to act as nuclei round which the water may deposit, air can be cooled considerably below its saturation point without depositing moisture. The reason for this phenomenon is that a very curved surface, such as that of a very minute drop of water, has, at a given temperature, a higher vapour pressure than a plane surface. Hence in the absence of dust a very minute drop, if it forms, will have a higher vapour tension than that ordinarily corresponding to the temperature, and hence will re-evaporate. If, however, the initial condensation can take place on a particle of dust, the water is spread out on a comparatively flat surface, and hence possesses the ordinary surface tension. As we shall see later, particles other than dust particles are capable of acting as nuclei for the condensation of water when saturated air is expanded.

In the case of some other substances, such as benzene, the adiabatic through A (Fig. 126) lies inside the saturation curve above and outside below. Hence in such a case adiabatic compression causes partial liquefaction, and the specific heat of the saturated vapour is positive.

The fact that if saturated steam is expanded adiabatically it becomes unsaturated, is the reason high-pressure steam escaping from a boiler does not burn the hand, while the steam which escapes from the spout of a kettle does burn. The steam as it escapes from the boiler expands almost adiabatically, and hence is unsaturated, and like a gas it only parts with its heat to a solid body slowly, so that the hand is not much heated. Saturated steam, on the other hand, on meeting the hand condenses to water at 100° C., and as it liberates its latent heat when condensing, a large amount of heat is communicated to the hand, and a burn results.

A consideration of the different slope of the isothermal and adiabatic

through a given point shown in Fig. 127 will show that if we compress a gas adiabatically the pressure will rise considerably faster than if we compress it isothermally. Thus if we start with air in the condition represented by the point A and compress it adiabatically till the volume is reduced to OM, the pressure will rise to MC, where CA is the adiabatic. The work done while compressing the air is represented by the area of the figure MCAN. If, however, the air is compressed at constant temperature along the isothermal AB, the work done is represented by the area MBAN. Suppose now a pump is used to compress air into a reservoir, unless the compression takes place very slowly the process will be approximately adiabatic, but the air stored in the receiver will gradually cool down to the atmospheric temperature, so that as far as the pressure in the receiver is concerned the air might have equally well have been compressed isothermally. Thus in practice the work performed by the pump represented by MCAN is greater than the energy finally stored in the compressed air. Next let us use the compressed air to work an engine, in which case again, unless the expansion is very slow, the change will be an adiabatic one, and the work done by the air will be represented by the area MBDN, where BD is the adiabatic through B. It will thus be seen that in a compressed air system of storing energy, since we can neither compress the air in the pump or expand it in the engine isothermally a considerable loss of power results, which is represented in the figure by the area BCAD.

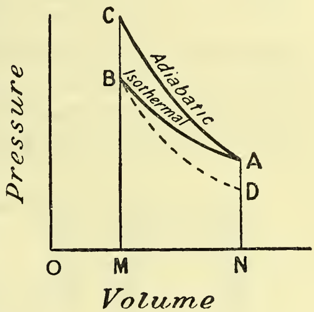


FIG. 127.

More work must be performed to compress a gas adiabatically than isothermally.

81. Heat Engines. Carnot Cycle. Second Law of Thermodynamics.

—In an ordinary condensing steam-engine the water is converted into steam in the boiler. This steam does external work as it expands in the cylinder, is then converted into water in the condenser, and finally pumped back into the boiler. Hence if no leakage occurred the same quantity of water, which is called the working substance, would be used indefinitely, and would be caused to go through the above cycle of operations again and again. The cycle of operations, however, which takes place in the steam-engine is somewhat complex, for the working substance not only loses heat owing to the external work performed and by communicating heat to the condenser, but also loses heat by conduction to the walls of the cylinder and to the air surrounding the pipes which convey the steam from the boiler

to the engine. Hence we shall first consider the working of an ideal engine in which the cycle of operations is very much simpler than that of any real engine. This cycle was first used by Carnot, and hence it is known as Carnot's cycle.

The Carnot cycle.

Suppose the working substance, say air, is contained within a cylinder W (Fig. 128), the walls and piston of which are perfect non-conductors for heat, and that the bottom of the cylinder is made of a perfect conductor of heat. Further, we will suppose that we have three stands, one, X,

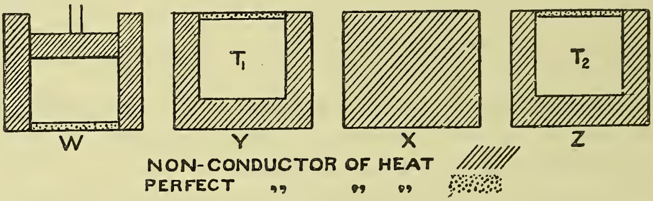


FIG. 128.

fitted with a perfect non-conducting top, and of the others one, Y, kept at a constant temperature T_1, and the other, Z, at a constant temperature T_2, these being each fitted with a perfectly-conducting top.

First place the cylinder on the stand Z, so that the working substance comes to a temperature T_2, the pressure and volume being as indicated by the point D (Fig. 129). Now place the cylinder on the non-conducting stand X, and increase the pressure, and so decrease the volume. The change must be adiabatic, since the non-conducting stand prevents the escape of the heat due to the compression. The compression must be stopped when the temperature has risen to T_1. The curve DA, which gives the relation between the pressure and volume during this operation, is an adiabatic. Next place the cylinder on the stand Y, and expand the substance by allowing the piston to rise till the volume and pressure are indicated by the point B. During this operation heat will flow into the working substance through the conducting bottom of the cylinder, so that the temperature will be constant throughout the process, and hence the curve AB will be a portion of the isothermal for the temperature T_1. Next remove the cylinder from Y, and place it on the non-

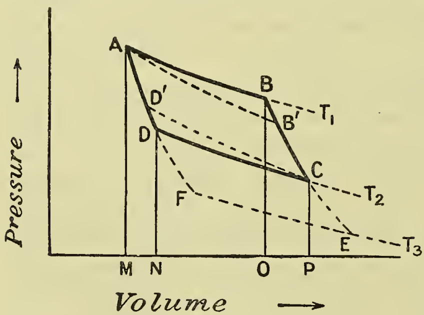


FIG. 129.

conducting stand X, and continue allowing the substance to expand till the temperature falls to T_2. This portion of the cycle, since the escape or supply of heat to the working substance is prevented by the non-conducting stand, is adiabatic. Next place the cylinder on Z, and force the piston down till the pressure and volume become the same as at the start. During this part of the operation heat is given out by the working substance to the stand, and CD is a portion of the isothermal for T_2. The cycle is now complete, the working substance having been returned to its original condition. While the working substance was being taken from A to B the external work done by the substance is represented by the area of the figure MABO. In the same way the work done while going from B to C is represented by the area of the figure OBCP. Hence the total work done *by* the working substance is represented by MABCP. Similarly the work done *on* the working substance by the external agent which compresses the substance along CD and DA is represented by the area of the figure PCDAM. Hence the net work performed by the working substance is represented by the area of the figure DABC, which represents the path along which the working substance has been taken during the cycle.

During the portion AB of the cycle heat has been supplied while the working substance had a temperature T_1, while during the part CD heat has been abstracted, the temperature of the working substance during the abstraction being T_2. If W is the work performed during the cycle, and H_1 and H_2 are the quantities of heat, measured in mechanical units, supplied along AB and abstracted along CD respectively, then since the final state of the working substance is exactly the same as at the start, so that the quantity of heat it contains must be the same at the beginning and end of the cycle, it follows from the first law of thermodynamics that the work done during the cycle must be equivalent to the heat which has been used. That is

$$W = H_1 - H_2$$

Next suppose that we start with the cylinder in contact with the stand Z and expand isothermally to C. Next place on X and compress adiabatically to B, that is till the temperature rises to T_1. Then place on Y and compress isothermally to A. Finally place on X and expand adiabatically to a temperature T_2, when the working substance will again be in its initial state as represented by the point D. In this cycle we have just reversed the various operations performed previously, and during the process a net amount of work represented by the area of the figure DCBA has been performed *by* the external forces, while an amount of heat H_2 has been absorbed at a temperature T_2, and an amount H_1

given out at a temperature T_1. Owing to the fact that it can be reversed, Carnot's cycle is called a reversible cycle.

It is important to clearly understand what is implied by a reversible cycle. When in the direct Carnot cycle the cylinder is on Y in order that heat may be communicated to the gas in the cylinder, the temperature of the gas must be *slightly lower* than the temperature of the reservoir Y. On the other hand, when the cycle is reversed the temperature of the gas must be slightly higher than that of the reservoir Y if heat is to flow from the working substance to the reservoir. However, if the expansion from A to B (Fig. 129) on the direct cycle and the compression from B to A on the reverse cycle is made very slowly, a very small difference in temperature will be sufficient to allow of the transfer of heat, and hence we are at liberty to assume that in the limit we could perform the operations which are postulated in the cycle and its reverse in which the temperature in the working substance is the *same* both in the direct and in the reverse operation along AB. A similar line of argument holds for the isothermal change along CD. If along, say, AB the gas is allowed to expand quickly, owing to the external work performed the gas will lose heat, and the heat will not be able to flow sufficiently quickly from the reservoir to keep the whole of the gas at a constant temperature, so that the portions of the gas in the cylinder more remote from the bottom of the cylinder which is in contact with the reservoir will be at a lower temperature than T_1, and hence the volume occupied by the gas will be less, so that the curve AB′ showing the relation between the pressure and volume during the change will slope from the point A more steeply than the isothermal AB, and hence less external work will be performed. In the same way during the change CD, if the compression is rapid the temperature of the parts of the gas remote from the bottom of the cylinder will be above T_2, and the curve CD′ showing the relation between pressure and volume will slope up more steeply than the isothermal. Hence the work done *on* the gas will be greater than when the change is very slow. It will thus be seen that the net work performed in the cycle, which is equal to the area of the figure AB′CD′, is less than before. We shall return to this matter later on.

Efficiency of a Carnot cycle.

In the Carnot cycle the amount of heat converted into work is $H_1 - H_2$, while from the first law the maximum amount of work which could be obtained from the heat supplied is H_1. The ratio of the amount of heat actually converted into work in the cycle to the heat supplied at the higher temperature is called the efficiency of the cycle. Hence the efficiency, n, of a Carnot cycle is given by

$$n = \frac{W}{H_1} = \frac{H_1 - H_2}{H_1} \quad . \quad . \quad . \quad . \quad (81)$$

If the lower temperature of a Carnot cycle were taken at a temperature T_3 lower than T_2, the work performed in the cycle would be represented by ABEF, which is greater than before. Hence the efficiency when the temperatures between which the cycle is taken is T_1 and T_3 is greater than that obtained when the limits of temperature are T_1 and T_2, so that the efficiency of any particular cycle depends on the temperatures at which heat is taken in and rejected by the working substance.

Attention must be drawn to the fact that in the case of all heat engines the heat converted into work $(H_1 - H_2)$ is always less than the heat H_1 supplied to the engine at the higher temperature. Thus the efficiency of all engines must be less than unity. Of course, as we have seen above, the efficiency is greater the lower the temperature of the cold reservoir, which we may by analogy with a steam-engine call the condenser, but it is not practicable to have a condenser at a lower temperature than that of surrounding objects, *i.e.* in the case of a marine engine at the temperature of the sea, so that practically a lower limit to the temperature range which can be employed is fixed. The upper limit allows us more freedom of choice, though here again mechanical considerations generally forbid very high temperatures being employed.

Efficiencies of all reversible cycles between the same temperatures are the same.

No engine can have a higher efficiency than a reversible engine working between the same two temperature limits, while so long as the cycle is *reversible*, the efficiency is independent of the kind of cycle or the nature of the working substance. For suppose we could have an engine A of which the efficiency N between two temperatures T_1 and T_2 is greater than the efficiency n of a reversible engine B working between the same two temperatures. Then let the heat taken in by A at T_1 be H_A, and that rejected at T_2 be H'_A, the work performed being W, so that $N = \frac{W}{H_A}$. Now suppose engine A coupled to engine B so as to make B work in the reverse direction, and let engine B take in a quantity of heat H'_B at the temperature T_2.

The quantity of heat H_B given out by B at the temperature T_1 is such that $n = \frac{H_B - H'_B}{H_B}$, or since the work done on this reversed engine is the work done by the other, so that $H_B - H'_B = W$, $n = \frac{W}{H_B}$. But by hypothesis N is greater than n, hence H_A must be less than H_B. Thus the reversible engine returns to the source of heat at the higher temperature more heat than is abstracted by the non-reversible engine, while the combination of the two engines is not supplied with mechanical energy from any outside source, the work done on B being entirely

derived from the work done by A. It also follows that H'_B is greater than H'_A, so that more heat is abstracted from the source at the lower temperature than is supplied to it. Thus the combination would be capable, without any outside assistance, of abstracting heat from a source at a given temperature and conveying this heat to a body at a higher temperature. Now this performance is as much against the result of all our experience as is the impossibility of producing perpetual motion. Hence we conclude that the supposition that any engine can be more efficient than a reversible one working between the same temperatures cannot be true. It follows, of course, that all reversible engines must have the same efficiency as long as they work between the same temperatures, and hence the efficiency must be independent of the nature of the working substance.

Second law of thermodynamics.

The statement made above that no engine or combination of engines or arrangement whatsoever can, unless supplied with energy from some outside source, convey heat from a body to another at a *higher* temperature forms what is known as the second law of thermodynamics.

Lord Kelvin has enunciated the second law in a slightly different form, namely: *It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of surrounding bodies.*

It must be carefully borne in mind that the law refers only to the work performed during a *cycle* of operations, in which the initial and final states of the working substance are exactly the same. Thus when a gas is allowed to expand against external pressure, it does work and becomes cooled, so that in this way it may do work although in the operation it becomes cooled below the temperature of surrounding objects. The final state of the gas is not, however, the same as the initial state, and if we attempt to bring the gas back into the initial state we shall find that the law holds.

We may also put the law in slightly different words, viz. that heat of itself never passes from one body to another at a higher temperature; and if by any means we cause heat to be transferred from a body to another at a higher temperature, we must in the process supply the system with energy from some outside source. Thus, when a reversible engine is worked backwards, heat is taken from the refrigerator and supplied to the source. During this operation, however, external energy has to be supplied to the engine, so that it is not working "by itself."

A Carnot's cycle being a thermal process which is independent of the nature of the substance in which the thermal changes take place, it at once becomes of interest to see whether we cannot utilise this fact in

order to define a scale of temperature independent of all properties of any particular kind of matter. The scales which we have used heretofore all depend on the change of some one physical property of some special kind of matter, thus on the increase in volume of mercury or hydrogen, the increase in resistance of a platinum wire, the thermo-electromotive force of a junction of two given metals, &c. A scale of temperature depending on Carnot's cycle and independent of the properties of any particular kind of matter has, however, been devised by Lord Kelvin, and to such a scale only can the title "absolute" be given with justice.

Kelvin's absolute scale of temperature.

If, as before, we imagine a Carnot's cycle in which a quantity of heat H_1 is drawn from a source at a temperature T_1, and an amount of work W is performed, H_2 units of heat being given out to the refrigerator at the temperature T_2, we may according to the first law measure H_2 and H_1 in terms of ergs, in which case $H_1 - H_2 = W$. If now, keeping H_1 and T_1 constant, we adjust the temperature T_2 so that the work done during the cycle is unity, then the two temperatures T_2 and T_1 will be such that if a Carnot's engine working between these temperatures takes H_1 ergs of heat from the source it will perform one erg of work. Next suppose that another cycle is taken, in which the lower temperature T_3 is so adjusted that when H_1 ergs of heat are drawn from the source at a temperature T_1, the work done in the cycle is two ergs. Then, according to Lord Kelvin, the difference of temperature between T_1 and T_3 is to be called twice the difference of temperature between T_1 and T_2. Proceeding in this way, we could define a series of equal temperature intervals, and thus form a thermometric scale. It will, however, be convenient not to call the interval $T_1 - T_2$, or $T_2 - T_3$, as above defined, one degree, since the scale thus constructed would not resemble the scale ordinarily employed. We will therefore suppose that T_1 is taken as the temperature of boiling water, and we will postulate that when H_1 units of heat are taken, by an engine working in a simple reversible cycle, from a source at the temperature of boiling water, and the refrigerator is at the temperature of melting ice, a hundred times the work will be done that would be done supposing the temperature of the refrigerator were one degree, on this new absolute scale, below the temperature of boiling water, and so on.

Let the lines T_0T_0 and $\text{T}_{10}\text{T}_{10}$ (Fig. 130) be the isothermals for the temperatures of melting ice and boiling water respectively, and let AB be an adiabatic cutting these isothermals at E and G. Suppose that if we go along the isothermal T_{10} from E to F an amount of heat H_{10} (measured in ergs) has to be supplied to the working substance to keep its temperature constant, and that through F we draw a second adiabatic CD cutting the isothermal T_0 at H. Then, if a simple reversible engine performs the

cycle EFHG, it will take in H_{10} units of heat at a temperature T_{10} and give out H_0 units of heat at a temperature T_0, while the work W_{10} done during the cycle will be represented by the area of the figure EFHG. Now draw nine isothermals between T_0 and T_{10}, so spaced that the area intercepted between any adjacent two and the two adiabatics is one-tenth of the area EFHG. Thus the area shown shaded is to be one-tenth of EFHG. Then the temperatures corresponding to these isothermals, if we call the temperature of melting ice 273° and that of boiling water 373°, are 283°, 293°, 303°, 313°, 323°, 333°, 343°, 353°, 363° on Lord Kelvin's absolute scale.

By the doctrine of the conservation of energy, the maximum amount of work we can possibly get from a quantity of heat H_1 is JH_1, if the quantity H_1 is expressed in calories, or simply H_1, if this quantity is expressed in ergs. Keeping the temperature T_1 of the source constant, the amount of work W obtained during a cycle will increase as the temperature of the refrigerator is lowered, until the temperature of the refrigerator becomes such that no heat is given to it during the compression portion of the cycle, the whole of the heat taken in being converted into work, so that $H_1 = W$, or the efficiency of the cycle becomes unity. We cannot imagine a refrigerator at a lower temperature than this, and hence may take it as the zero on this new absolute scale. It is found that the zero thus defined coincides with the absolute zero as given by a perfect gas, and that the new absolute scale agrees very nearly with that of a gas thermometer containing a perfect gas. So that the use of the thermometric scale derived from the expansion of a perfect gas is justified.

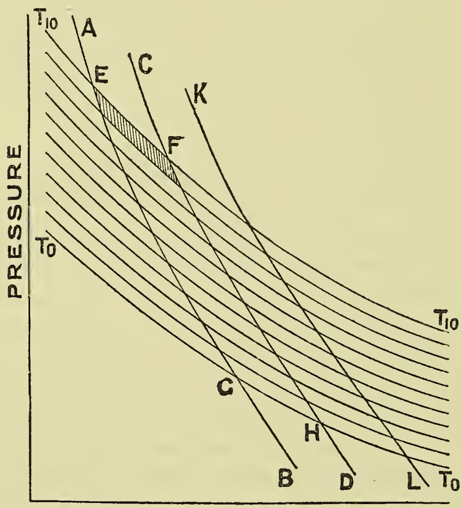


FIG. 130.

If, in Fig. 130, any other adiabatic KL is drawn, then this, together with either of the others, will cut off equal areas between consecutive isothermals. Thus the area intercepted by any two adiabatics and any two isothermals T_1 and T_2, say, will be proportional to the difference of temperature $T_1 - T_2$, for each degree of this difference will correspond to

an equal small area k, such as the one shown shaded. Thus we shall have

$$W = k(T_1 - T_2),$$

where k is a constant depending on the two adiabatics taken. Since $W = H_1 - H_2$, this gives

$$H_1 - H_2 = k(T_1 - T_2).$$

Now if we make T_2 the absolute zero, there will be T_1, small areas each equal to k included in the cycle, and H_2 will be zero, so that in this case

$$W = H_1 = kT_1.$$

Now the efficiency of a reversible cycle is given by

$$n = \frac{H_1 - H_2}{H_1}$$

Hence, substituting for $H_1 - H_2$ and H_1,

$$n = \frac{H_1 - H_2}{H_1} = \frac{k(T_1 - T_2)}{kT_1} = \frac{T_1 - T_2}{T_1} \quad . \quad . \quad (a)$$

or since $W = H_1 - H_2$, we have

$$\frac{W}{H_1} = \frac{T_1 - T_2}{T_1} \quad . \quad . \quad . \quad . \quad . \quad (82)$$

Equation (a) may be written

$$1 - \frac{H_2}{H_1} = 1 - \frac{T_2}{T_1}$$

so that

$$\frac{H_1}{H_2} = \frac{T_1}{T_2} \quad . \quad . \quad . \quad . \quad . \quad (83)$$

or the ratio of the heat taken from the source by a reversible engine to the heat given up to the refrigerator is the same as the ratio of the temperature, on the absolute scale, of the source to that of the refrigerator.

82. Reversible and Irreversible Processes.—We have already in the preceding section considered the condition that the isothermal expansion of a gas is reversible, namely, that it should be conducted so slowly that at no time should there be an appreciable difference of temperature between different parts of the gas. If the expansion is so rapid that differences of temperature do occur, then when we try to reverse the operation by compressing the gas we do not exactly retrace the path, and hence the operation is irreversible. Again, suppose that the base of the cylinder through which the heat has to pass on its way from and to the receiver is not a perfect conductor of heat. In order to cause a flow of heat the temperature on one side of the base would then have

to be greater than that on the other. Thus when the gas was receiving heat it would have to be at a lower temperature than that of the receiver, while when the gas was giving up heat its temperature would have to be higher than that of the receiver, so that the process would be irreversible. Finally, suppose that the piston which confines the working substance in the cylinder moves with a certain amount of friction. Thus some of the work performed by the working substance when expanding will be employed in overcoming this friction, so that the external work performed will be reduced. On the reverse process, however, the external agent will not only have to do the work of compressing the working substance, but will also have to do work against the friction. Hence owing to friction the external work obtained during expansion is less than the work which has to be performed on the working substance during compression, and the process is not reversible.

Thus we see that a cycle in which we have (*a*) different parts of the working substance at different temperatures at the same instant, or (*b*) the passage of heat by conduction, or (*c*) the production of friction is not reversible. In practice, since it is quite impossible to entirely eliminate the above effects, no cycle can be absolutely reversible. Nevertheless, since in some processes, such as those in the Carnot cycle, where the effects due to these causes can, at any rate in imagination, be reduced to an infinitesimal amount, we are justified in speaking of such processes as reversible. There are other processes, however, which are fundamentally irreversible. Thus when heat passes by conduction, say along a copper rod, from a hot body to a cold, or when mechanical energy is converted into heat by friction, no known method enables us to reverse the operation, for heat cannot be made to flow along the rod from the cold body to the hot, neither can we by friction convert heat into mechanical energy. It is to be noted that heat can pass from a hot body to a cold without performing any work, as happens when conduction takes place; but to transfer heat from a cold body to a hot, as by means of a reversed Carnot engine, a certain amount of mechanical energy has to be supplied from some outside source.

83. Actual Heat Engines and their Efficiency.—A diagrammatic sketch of a steam-engine is shown in Fig. 131. Steam from the boiler A passes through the pipe B and a valve E to the cylinder D. A second valve E and a pipe F communicates with a condenser G, where the steam passes through a series of pipes which are kept cool by a circulation of cold water. The water formed by the condensation is forced by the pump H back into the boiler. When the piston is at the top of the stroke the valve C is opened, and steam from the boiler enters the cylinder and drives the piston down. After the piston has descended through a portion of its stroke, the valve C is closed so that no

more steam can enter the cylinder. The point on the stroke of the piston where the valve closes is called the cut-off, and the pressure and the travel of the piston is represented by the portion BC (Fig. 19) of the indicator diagram (see § 15). After the close of the valve C the steam in the cylinder continues to drive the piston down, and expands approximately adiabatically, the change being represented by CD, till at the end of the stroke the valve E is opened. The steam then passes into the condensers, in which a fairly good vacuum exists, and is condensed. The valve E remains open till the piston reaches nearly to the top of the stroke, when it closes and the small amount of steam remaining in the cylinder is then compressed by the piston so that at the top of the stroke the pressure has almost risen to that existing in the boiler. The valve C is then again opened and the whole process repeated.

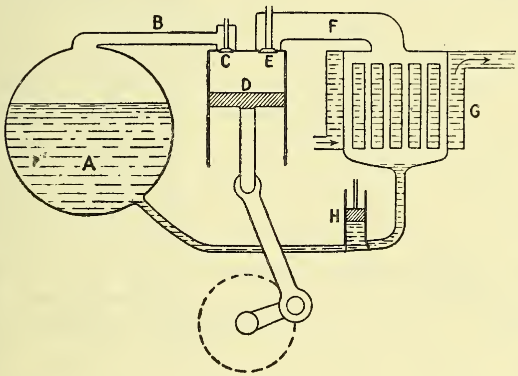


FIG. 131.

The working substance (water) takes in heat at the temperature existing in the boiler, and gives out heat at the temperature existing in the condenser, but the operations which occur are by no means reversible. Thus a certain amount of heat is lost by conduction from the outside of the pipe B, while a further loss is due to communication of heat to the piston and cylinder walls, and, as we have seen, whenever such conduction of heat occurs the process is irreversible. Another effect which has a considerable influence on the efficiency of the steam-engine is due to condensation of steam into water *inside* the cylinder. The steam as it enters the cylinder is saturated, and, as we have seen in § 80, when saturated steam is expanded adiabatically some of the steam becomes water, so that during the part of the stroke represented by CD (Fig. 19) on the indicator diagram the cylinder is filled with steam charged with water, forming a sort of cloud. A certain amount of this condensed steam has time to precipitate on the walls of the cylinder.

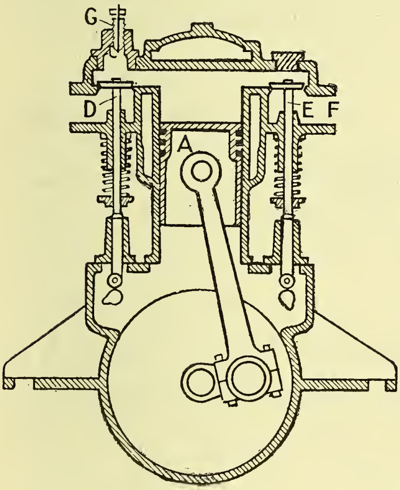
Again, when the steam enters it is at a higher temperature than that of the walls of the cylinder, and hence some condenses on the walls, forming water. Owing to this condensation and the latent heat thus liberated the walls of the cylinder are heated. When, however, the cylinder is in communication with the condenser the pressure is so low that the water on the hot walls of the cylinder evaporates, and hence since its latent heat has to be supplied by the cylinder walls, these walls are cooled, and thus when fresh steam enters during the next stroke steam is again condensed. This process of alternate condensation of hot steam on the walls and subsequent re-evaporation is irreversible, and the heat which is carried by this steam from the boiler is in a great measure wasted.

Thus if the condensation of steam could be prevented a considerable economy could be effected. There are various methods by which such condensation can be reduced. Thus the cylinder walls can be kept hot by surrounding them by a jacket which is supplied with steam from the boiler. Of course a certain amount of heat is thus wasted in the jacket, but experiment has shown that the amount thus wasted is not nearly as great as that wasted when condensation takes place in the cylinder. Another method of reducing condensation is to heat the incoming steam to a higher temperature than that existing in the boiler, so that the saturated steam which leaves the boiler is converted by this superheating into unsaturated steam, and such unsaturated steam can be cooled to a certain extent before it becomes saturated, that is before it can condense. Another method employed is to perform the expansion of the steam in several cylinders. Thus the difference in temperature between that at which the steam enters and leaves any particular cylinder is reduced, and hence the *variation* in the temperature of the cylinder walls is reduced. Engines in which the expansion takes place in stages in this way are called compound steam-engines, and are known as double, triple, or quadruple expansion, according as there are two, three, or four stages in the expansion. In a triple expansion engine with steam-jacketed cylinders supplied with saturated steam at 190° C., the condensed water being at 40° C., the efficiency is about ·17. That is, the ratio of the heat required to raise a pound of water from 40° C. and to convert it into saturated steam at 190° C. to the indicated work performed by this engine by the consumption of one pound of steam is ·17. If we take the boiler into account and consider the heat developed by the fuel in the production of one pound of steam, the efficiency will be only about ·12. A reversible engine working between the same temperature limits (190° C. and 40° C.) would have an efficiency (§ 81) of ·32. It will thus be seen that the efficiency of the actual steam-engine is only about half that of the ideal reversible engine.

The case considered above is that of an exceptionally efficient steam-engine. An ordinary locomotive is only about a third as efficient.

The internal-combustion engine.

In the case of a gas or petrol engine the heat of combustion of the fuel is actually liberated inside the cylinder of the engine, in place of being liberated in a furnace, and thence conducted to water contained in a boiler, this water being employed to convey the heat to the cylinder. A section of a petrol engine is given in Fig. 132 (*a*). In the ordinary cycle of operations employed, starting with the piston A at the top of the stroke, as the piston descends the valve D is opened and a mixture of petrol vapour and air is drawn into the cylinder. The line *abc* on the indicator



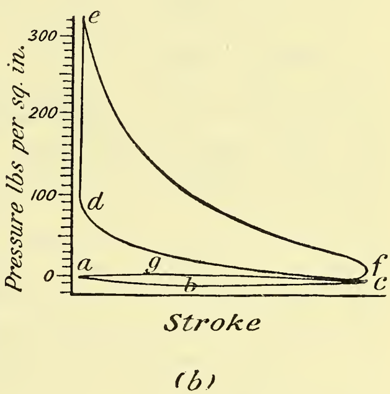


FIG. 132.

diagram (Fig. 132 (*b*)) represents this induction stroke. The valve D is then closed, and as the piston returns the charge is compressed almost adiabatically, as represented by the line *cd*. When the piston is at the top of the stroke a spark is passed at the plug G which fires the charge. Owing to the heat liberated by the combustion the pressure rises very rapidly, as shown by the line *de*. During the next outward stroke the gases expand and do work on the piston. The expansion is almost adiabatic, and is represented by the line *ef*. When the piston is near the end of the stroke the valve E opens and the gases escape through the tube F. On the return stroke of the piston the products of combustion are further expelled, the valve E remaining open, and the remaining portion of the cycle is represented by the line *fga*. The work done by

the charge during the whole cycle is represented by the area of the figure *efd*, and it will be observed that only a single working stroke is obtained for two complete revolutions of the crank-shaft, that is, for four to or fro strokes of the piston. Hence the expression four-cycle engine.

The efficiency of a large gas-engine may reach the value ·37, which, it will be observed, is considerably higher than that of the very best steam-engine. Even with such small engines as are used in motor-cars, the efficiency attained may be as high as ·28. The chief reason for these comparatively high efficiencies is the high temperature at which the heat is communicated to the working substance, although the fact that the working substance rejects heat also at a high temperature detracts from the efficiency, it being impracticable to work with a *lower* value for this temperature than about 800° C.

In a steam-turbine the steam from the boiler is allowed to expand by escaping through a channel of gradually increasing cross-section. As the steam expands it does work, and this work is employed in increasing the velocity of the steam along the channel, that is, increasing the kinetic energy of the steam. The rapidly flowing steam is then allowed to impinge on a series of curved blades, which are fixed to the circumference of a wheel. Owing to the action of the blades the direction of motion of the steam is altered, and momentum is communicated to the wheel, the action being similar to that considered in § 51. The steam-turbine gets over the difficulty due to steam condensation in the cylinder of the reciprocating engine, but introduces difficulties of its own. Thus the velocity of the steam before it impinges on the wheel is very great, and since the linear speed of the vanes must not differ greatly from half the speed of the steam if the maximum amount of momentum is to be communicated to the wheel, this latter has to rotate at very high speeds indeed. Further, there is a considerable amount of friction between the steam and the surfaces which guide its direction of motion, as also between the rapidly rotating wheel and the steam that fills the space in which it rotates. As a result the efficiency attained by steam-turbines is not very much greater than that attained by the best type of reciprocating engine.

BOOK III

PERIODIC MOTION AND SOUND

PART I

PERIODIC MOTION. WAVE MOTION

CHAPTER I

VIBRATORY MOTION

84. Simple Harmonic Motion.—Suppose we have (Fig. 133) a point Q which moves round a circle MANB with a uniform speed, and from the position occupied by Q at any instant we draw a perpendicular QP on the diameter MON of the circle. Then as Q revolves in the circle the foot P of the perpendicular will move backwards and forwards along the line MN, and the point P is said to have a *simple harmonic motion* along the line MN. The letters S.H.M. are generally

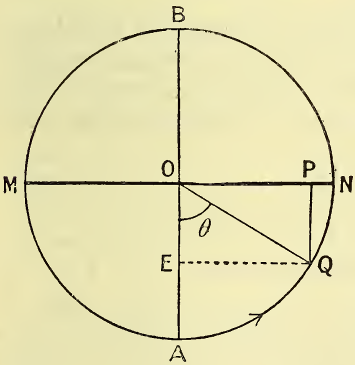


FIG. 133.

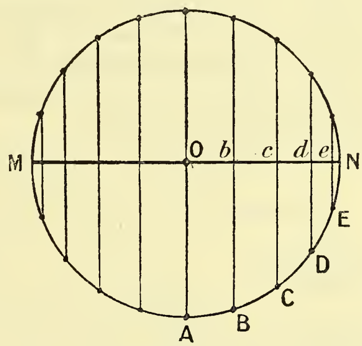


FIG. 134.

used as an abbreviation for simple harmonic motion. It must be borne in mind that although we have used the point Q moving at a uniform speed in a circle called the circle of reference, on MN as diameter, to *define* what we mean by a simple harmonic motion along MN, yet the point P may execute a S.H.M. without there being a corresponding point moving in the circle.

If we take a number of points in the circle of reference at equal distances round the circumference, and as in Fig. 134 draw perpen-

diculars through each point, then the particle P which is executing the S.H.M. will describe the space between the feet of two consecutive perpendiculars in equal times. Thus each of the spaces ob, bc, cd, de, and eN is described in an equal time, namely, a twentieth of the time the particle moving in the circle of reference takes to complete one whole revolution. Thus starting with the point P when it passes through the centre O moving towards the right, the speed gradually decreases till the point N is reached, the movement at the end of the path being comparatively slow. The direction of motion is then reversed and the speed increases till the centre O is reached. As the point P moves away from O towards the left the speed gradually decreases till the point M is reached, when the motion is reversed in direction, the speed again increasing as P moves towards the centre.

Periodic time of S.H.M.

The time the point P takes to move from O to N, from N to M, and then from M to O is called the *periodic time* or *period* of the S.H.M. Since the periodic time is also the time the point in the circle of reference takes to complete a whole revolution, it follows that the periodic time is the interval between two consecutive passages *in the same direction* of the point P through any given point in its path. The number of complete to and fro passages of P in a second is called the *frequency*.

If o is the angular velocity (§ 19) of the point on the circle of reference, we have that the time occupied in completing a revolution is $2\pi/o$. Hence the period T is connected with the angular velocity of the point in the circle of reference by the relation $oT = 2\pi$. If N is the frequency, we have

$$N = 1/T = o/2\pi \quad . \quad . \quad . \quad . \quad (84)$$

Amplitude of S.H.M.

The maximum distance of the point P from the mid point of its path, that is OM or ON, is called the *amplitude* of the S.H.M.

If we count the time t from the instant when the particle P passes through O when moving from left to right, and the angle made by the corresponding particle Q in the circle of reference with the perpendicular OA is θ, we have (Fig. 133)

$$OP = OQ \sin \theta.$$

But $\theta = ot$, where o is the angular velocity of Q, and $OQ = ON =$ the amplitude, a. Hence calling x the distance OP between P and the centre O,

$$x = a \sin ot$$

$$= a \sin \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad (85)$$

The angle θ is called the *phase* at the given time t of the particle which is executing the S.H.M. It is evident that the displacement of the point P which is executing the S.H.M. is equal to the resolved part parallel to MN of the displacement of the particle Q in the circle of reference. Hence the velocity and acceleration of P will be the resolved parts of the velocity and acceleration of Q. Now the speed of Q is equal to oa and is along the tangent QT (Fig. 135), and the resolved part of this velocity parallel to MN is $oa \cos \theta$, for the angle CQT is equal to the angle AOQ. Hence the velocity, v, of P is $oa \cos \theta$, or since $\theta = ot$

Phase of a S.H.M.

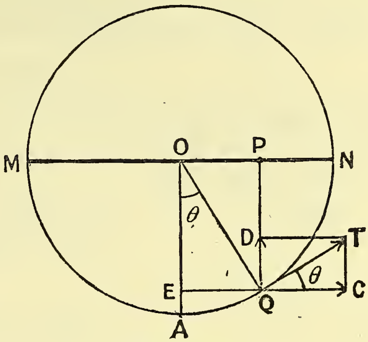


FIG. 135.

$$v = oa \cos ot \quad . \quad (86)$$

If x is the displacement OP at the time t, $\cos \theta = \frac{\sqrt{(a^2 - x^2)}}{a}$.

Thus

$$v = o\sqrt{(a^2 - x^2)} \quad . \quad . \quad . \quad . \quad (87)$$

since the acceleration of a point moving in a circle with constant angular velocity o is equal to ao^2 and is directed along the radius (§ 6). The acceleration a of P is therefore the resolved part parallel to MN of an acceleration ao^2 along QO. Since we have taken the positive direction for x and the velocity to be towards the right, while the resolved part of QO is towards the left, the acceleration, f, must be negative for the position of P shown in the figure. Hence

$$f = -ao^2 \sin \theta = -o^2 x \quad . \quad . \quad . \quad (88)$$

When the displacement is negative, *i.e.* when P is to the left of O, the acceleration is positive, that is, it is towards the right. Since O is constant it follows that in a S.H.M. the acceleration at any point is proportional to the displacement and always acts towards the centre O.

The angular velocity o of the point in the circle of reference is connected with the periodic time T by the relation $o = 2\pi/T$, and as we have seen above,

$$o^2 = -\frac{f}{x}$$

Hence

$$T = 2\pi\sqrt{-\frac{x}{f}} \quad . \quad . \quad . \quad (89)$$

It must be remembered that as f and x always differ in sign, the quantity under the square root is always positive.

We have now to consider in what manner a force must act on a particle of mass m to cause it to execute a S.H.M. Since the force is always equal to the mass multiplied by the acceleration (§ 8), we see that the force must always act towards the middle point O of the path along which the S.H.M. takes place, and when the displacement of the particle is x the magnitude of the force must be given by

$$F = -o^2 mx \quad . \quad . \quad . \quad . \quad (90)$$

The negative sign indicates that when the particle is displaced to the left of O, *i.e.* in the positive direction, F must act in the negative direction, *i.e.* from right to left.

Since $$o = 2\pi/T$$

$$F = -\frac{4\pi^2 mx}{T^2}$$

or $$T^2 = -4\pi^2 m\frac{x}{F} \quad . \quad . \quad . \quad . \quad (91)$$

Since x and F always differ in sign, the right-hand side of this equation is always positive.

As $4\pi^2 m$ is a constant, and the periodic time T is a constant, it follows that F/x is also constant. That is, in a S.H.M. the ratio of the force acting on a particle which is executing a S.H.M. when the displacement is x to that displacement is a constant. Hence the converse is also true, namely, if we have a particle which is so situated that when displaced from its position of rest a force is called into play which tends to bring the particle back towards its undisplaced position and is proportional to the displacement, then this particle will execute a S.H.M.

If the force of restitution[1] called into play by *unit* displacement is r,

$$T = 2\pi\sqrt{\frac{m}{r}} \quad . \quad . \quad . \quad . \quad (92)$$

So that the periodic time T is equal to 2π times the square root of the ratio of the mass of the particle to the force of restitution when the displacement is unity.

Angular S.H.M.

We have hitherto supposed that the particle which executes the S.H.M. moves backwards and forwards along a straight line. Next suppose a particle P moves backwards and forwards along an arc of a circle MON (Fig. 136), and the angle c the radius through the position of P at a time t makes with

[1] This force will be of the opposite sign to x.

the radius CO through the mid position of the path satisfies the equation

$$c = \theta \sin \frac{2\pi t}{T}$$

where θ is the angle OCN and T is the time the particle P takes to complete a whole cycle. Then the particle P is said to execute an angular simple harmonic motion, of amplitude θ and period T. If the radius of the circle becomes very great the arc MON becomes a straight line and P executes an ordinary S.H.M.

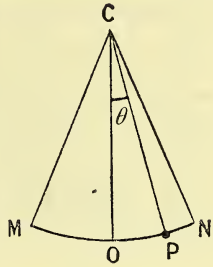


FIG. 136.

Similarly, if a body vibrates about an axis in such a way that the angle through which it is turned from its mean position at any time satisfies the above relation, the body is said to execute an angular harmonic motion.

In case of angular S.H.M., equation (89), p. 215, takes the form

$$T = 2\pi\sqrt{-\frac{\theta}{f'}} \quad . \quad . \quad . \quad . \quad (93)$$

where f' is the angular acceleration when the angular displacement is θ. Similarly, if when the angular displacement is θ the restoring couple is C, and I is the moment of inertia of the system about the axis, equation (91) becomes

$$T^2 = -4\pi^2 I \frac{\theta}{C} \quad . \quad . \quad . \quad . \quad (94)$$

A proof of these relations may be left as an exercise for the student.

The discussion of many problems involving S.H.M. is simplified by drawing a curve, called a harmonic curve, such that the abscissæ are times and the ordinates the corresponding displacements from the median position of the particle. This median position is often called the position of rest or the equilibrium position. Such a curve is given in Fig. 137, where a point such as p' on the curve OaBcD represents the displacement of the particle at the time OP′. The curve can easily be drawn by means of the circle of reference NQPM, for if the circumference be divided into a number of equal segments, each of these segments will correspond to equal times, *i.e.* to equal increases in the x of the harmonic curve, and the distance between the centre O′ and the feet of the perpendiculars from the points on the circle to NM will give the corresponding displacements.

Harmonic curve, or sine curve.

Since in a S.H.M. the displacement at a time t is equal to a sin ot, if the abscissa of a point on the harmonic curve is x, so that $x = t$, and the

corresponding ordinate, that is, the displacement of the particle is y, we have

$$y = a \sin ox \quad . \quad . \quad . \quad . \quad (95)$$

That is, the ordinate of the harmonic curve is proportional to the sine of an angle which is itself proportional to the abscissa, and for this reason the harmonic curve is often called the *curve of sines* or simply a *sine curve.*

The actual form of the curve depends on the amplitude a and the periodic time T, for the maximum ordinate Aa or Cc (Fig. 137) is a, while OD is equal to T, for OD represents the time occupied by the particle in performing a complete vibration.

85. Composition and Resolution of S.H.Ms.—Suppose we have two S.H.Ms. of the same period and amplitude but in directions at right angles to one another, and we require to find the resultant motion.

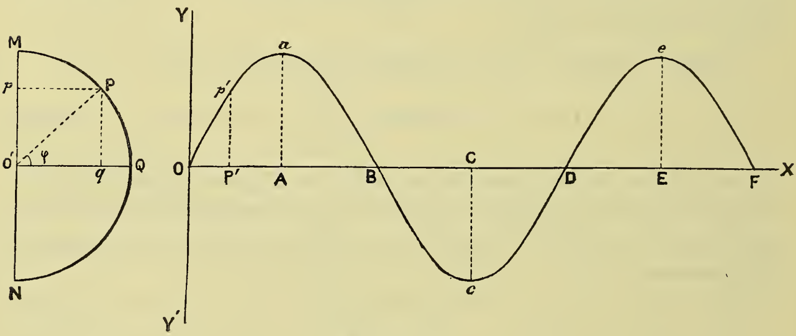


FIG. 137.

Let one motion take place along KL (Fig. 138) and the other along MN, and let KMLN be the common circle of reference. Divide the circumference of this circle into an even number of equal parts, say twelve, and through these points draw lines parallel to KL and MN as in Fig 138. If the S.H.Ms. are in the same phase, the extreme positive elongation will occur at the same instant in each. Hence if, as is usual, we consider from O to L to be the positive direction for one motion, and from O to M that for the other, L will be the position of maximum positive elongation for one and M for the other.

Starting then from the instant when both the motions are passing through the position of rest O, and the positive displacement is increasing, the points a, b, M will represent the displacements at times $T/12$, $2T/12$, and $3T/12$ due to S.H.M. along NM, while the points c, d, L will represent the displacements at the same instants due to the motion along KL. Hence the actual positions of a particle which is moving with the two

S.H.Ms. will be O, e, f, Q, &c. The resultant motion is therefore along the straight line QQ′ which is inclined at 45° to the directions of the two S.H.Ms. Hence the resultant displacement along QQ′ can be represented by

$$R = \sqrt{2}.a \sin ot.$$

This represents a S.H.M. of which the periodic time is the same as that of the two components (o being the same for all three), and of which the amplitude is $\sqrt{2}.a$.

If the two S.H.Ms., instead of being in the same phase, differ in phase by half a period, or 180°, then the resultant motion will be a S.H.M. along PP′ of amplitude $\sqrt{2}.a$.

If the two components differ in phase by 90°, or a quarter period, when one S.H.M. is at its extreme elongation the other will be passing

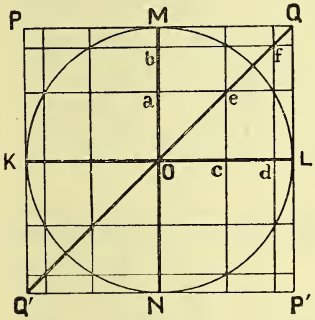


FIG. 138.

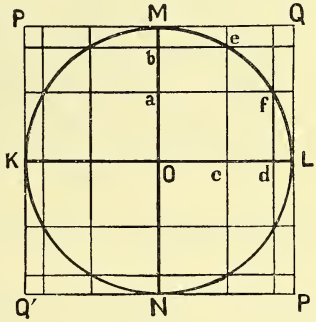


FIG. 139.

through its position of rest. Suppose that when the moving particle is at the point of extreme positive elongation (M), Fig. 139, as far as its motion along MN is concerned it is passing through O from left to right, owing to the motion along KL. Then at successive intervals of $T/12$ it will be displaced to b, a, O respectively with reference to one motion, and to c, d, L with reference to the other; and hence its resultant position will be e, f, L, &c. The resultant motion will thus be uniform motion in the circle of reference in the direction MLNK. If, however, when the particle is displaced to L by the horizontal motion, it is passing through O in the direction NOM, the resultant motion will be in the circle of reference but in the direction LMKN. Thus uniform motion in a circle may be regarded as the resultant of two S.H.Ms. of equal amplitude and period, but differing in phase by 90° and at right angles to one another.

For any other difference of phase the resultant motion will be in an ellipse, which will touch all four sides of the square PQP′Q′.

When either the amplitudes or periods of the two component S.H.Ms. are different, we cannot use the same circle of reference for the two motions. Suppose the period of the vertical S.H.M. is 2/3 that of the horizontal S.H.M., and the amplitude of the vertical motion is 2/3 that of the horizontal, the phases being the same. Let M′AN′ (Fig. 140) and K′BL′ be the two circles of reference, the diameter M′N′ being 2/3 the diameter K′L′, since the amplitude of the motion along MN is 2/3 of that along KL. The circumferences of these two circles must next be divided into parts that are traversed by the tracing-points in equal times. It is convenient to divide the quadrant into a whole number of parts in each case, hence in the above example we divide the circle M′AN′ into eight parts, and the circle K′BL′ into twelve parts. The period of the motion along MN being 2/3 of that along KL, the tracing-point of the

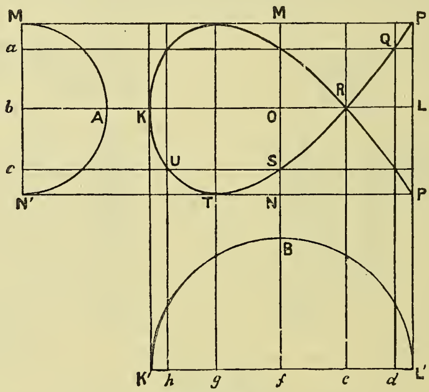


FIG. 140.

circle M′AN′ will traverse the circumference, while the tracing-point of the circle K′BL′ traverses 2/3 of the circumference. Thus the tracing-point in M′AN′ will traverse 1/8 of the circumference in the same time that the tracing-point in K′BL′ traverses 1/8 of 2/3 or 1/12 of the circumference; and hence we have divided the circles so that the arcs will be traversed in equal times. The phase of the motions being the same, the two extreme positive elongations occur simultaneously, and the particle starts at P. At the end of the interval chosen for subdividing the circles it has moved down to *a*, and horizontally to *d*, and hence its position is at Q. At the end of the next interval it has moved downwards to *b*, and horizontally to *e*; it is therefore at R. Similarly it travels to S and T. At T the particle has reached its extreme elongation in the vertical direction, and hence it now begins to move upwards, and during the next interval it reaches *c*. It continues, however, to

move to the left in the horizontal direction, and at the end of the interval is displaced to h. The actual position is thus U. In a similar manner the position at the ends of the remaining intervals can be found, and the path will be given by the line PRTKRP′. When the particle reaches the point P′, which it does after one complete period of the slower vibration (*i.e.* the horizontal) and one and a half periods of the faster, it will retrace the path, returning to P after two complete periods of the slower S.H.M. and three of the faster.

If the phases of the two components are not the same, the resultant motion would be different; the method of drawing the curves is, however, the same as in the above example. Some of the figures obtained are given in Fig. 141, where the phase of the vertical S.H.M. is increased by 45° between each figure and the next. In Fig. 142 another series of curves is given, in which the periods of the component S.H.Ms. are as 1 : 2, the amplitudes being the same. In this case the

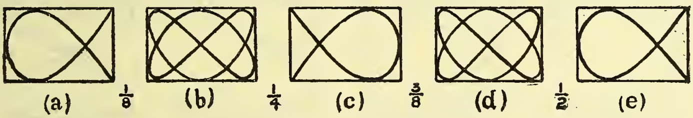


FIG. 141.

phase of the S.H.M. of shorter period is advanced by 30° between each figure and the next.

The above are all examples of the composition of two S.H.Ms., the periods of which are commensurate; that is, the ratio of the periods is expressed by *simple* whole numbers, so that, after a comparatively short time, equal to the least common measure of the periodic times, the particle will come back to its starting-point and the curve will then be retraced. If, however, the periods are not commensurable, the particle will not come back to its starting-point till after an infinite number of complete periods; that is, not at all.

There is one case which is of considerable interest, namely, when the periods can very nearly be represented by two simple whole numbers. If, for instance, the periods are as 2 : 1, then, as we have seen, we get a series of curves according to the difference in phase between the component motions; in each case, however, the curve is constant in form. Suppose now the periods, instead of being as 2 : 1, are as 201 : 100, and that the S.H.Ms. start in the same phase, then the path of the particle will be very nearly like (a), Fig. 142. However, when the slower S.H.M. has completed one vibration, the other, instead of having exactly completed two vibrations, will have completed two whole vibrations, together with 1/100 of another; it will thus have gained in phase on the other by

1/100 of a period or $360/100 = 3^\circ{\cdot}6$. This gain will continue till after eight periods of the slower vibrations, the difference in phase will amount to $28^\circ{\cdot}8$, and hence the curve traced out will resemble (*b*), Fig. 142. The difference in phase will continue to increase, and so by a continuous modification the curve will pass in succession through all the forms shown in Fig 142, first from (*a*) to (*g*), and then back from (*g*) to (*a*). For after 100 periods of the slower vibration, the quicker will have made a whole vibration more than it would have made if the ratio of the periods had been exactly 2 : 1, and for an instant the curve will again take the form of (*a*), Fig. 142, and will then go through the whole series again. If then we observe the interval which elapses between two consecutive passages of the curve through any particular form, say (*a*), Fig. 142, we can infer that in this interval one of the S.H.Ms. has exactly gained one vibration on the other.

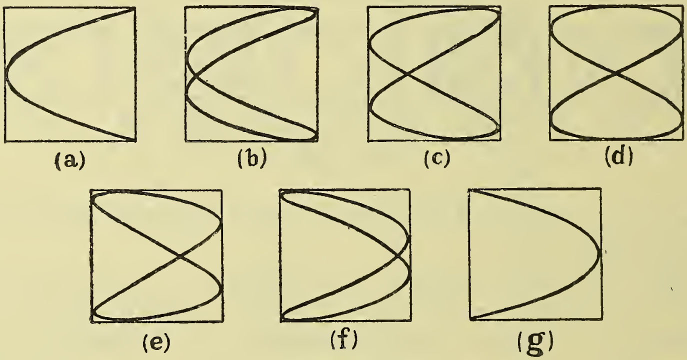


FIG. 142.

If the S.H.Ms. which have to be compounded are in the same direction, the simplest method of performing the composition is by means of the corresponding harmonic curves.

Let ABCDE (Fig. 143) be the harmonic curve corresponding to one S.H.M., so that AM represents the amplitude and BN the period, and let *abcde* represent another S.H.M. of amplitude *a*M and period *bf*, which starts in the same phase as the other. Then the resultant displacement will be obtained by adding together the displacements due to the two S.H.Ms. Thus at a time represented by the point L the total displacement will be equal to PL + *p*L, while at a time represented by K, the component displacements being in opposite directions, the total displacement is equal to KQ − K*q*, and since KQ is equal to K*q*, the displacement is zero. Hence if we construct a curve such that the ordinates are everywhere equal to the algebraic sum of the ordinates of the two component

curves, this curve will represent the resultant displacement. The resultant thus obtained is shown dotted in Fig. 143.

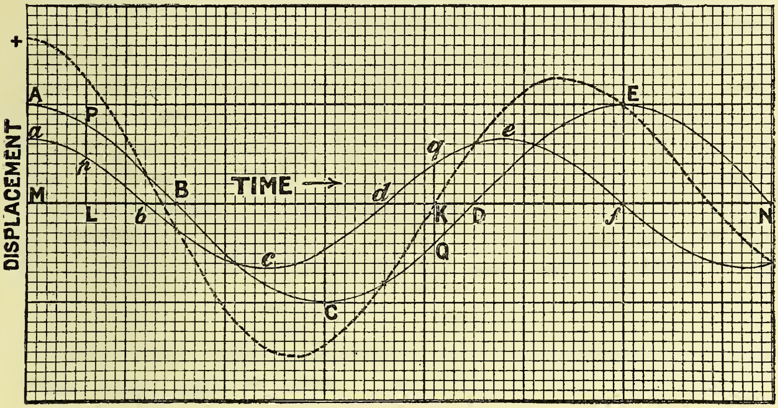


FIG. 43.

In Fig. 144 the same curves are compounded, but the time scale is made smaller, so that more periods of each curve may be shown. It will be seen that the resultant curve, although not a sine curve, is a periodic curve, and hence the resultant motion is periodic, the period being equal

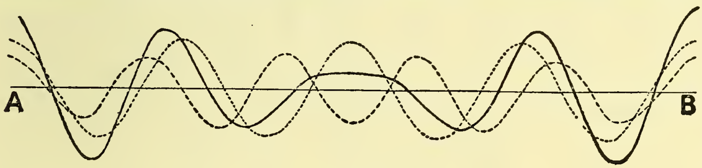


FIG. 144.

to AB, *i.e.* to five times the period of the quicker vibration, or four times that of the slower.

If the two S.H.Ms. to be compounded are of nearly the same period, say in the ratio of 9 : 10, then the compound harmonic curve obtained

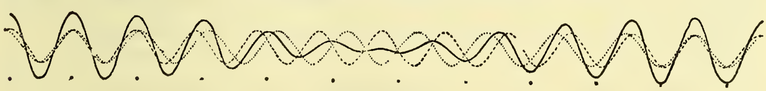


FIG. 145.

will, as is shown in Fig. 145, everywhere approximate to the form of a sine curve, but the amplitude will alternately wax and wane; the maxima occurring when the component vibrations are exactly in phase, and the minima when the phases differ by half a period. As in 9 periods of the

slower vibration there occur 10 periods of the quicker, in this interval one will have gained exactly one period on the other, and they will again be in the same phase. Thus the curve shows that the maxima occur at every 10th period of the quicker vibration.

When we have an alternate waxing and waning of the resultant motion due to two S.H.Ms. of nearly the same period being compounded, the phenomenon is generally referred to as a case of the production of *beats* (§ 100). Not only can we compound together two or more S.H.Ms. to give a resultant periodic motion which is not necessarily a simple harmonic one, but we may resolve a periodic motion into component S.H.Ms., and Fourier first showed that *any* periodic motion *whatever* can be considered as the resultant of a number of S.H.Ms.

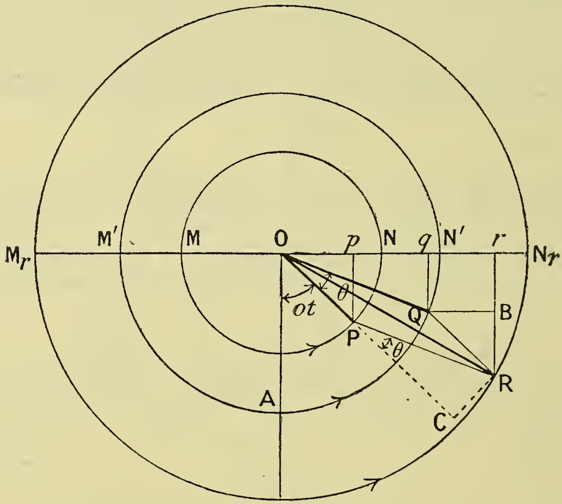


Fig. 146.

When S.H.Ms. in the same direction but of *different periods* have to be compounded, the method described above is most convenient. In many cases which occur in practice, however, the component S.H.Ms. have all the same period, when another method is often preferable.

Suppose we require to find the resultant of two S.H.Ms. of which the amplitudes are a_1 and a_2, and which have the same period T, but the phase of the first is θ behind that of the other. Draw the two circles of reference, MPN and M′QN′ (Fig. 146), of radii a_1 and a_2 respectively, and let P be the position of the point in the circle of reference MPN at a time t, so that the angle AOP is equal to ot. Then if OQ is drawn so that the angle QOP is θ, Q will represent the position of the point in the circle of

reference M′AN′ at the time t. On OP and OQ as adjacent sides complete the parallelogram OPRQ and join OR. Then the displacement of the particle at the given instant is Op due to the first S.H.M., and Oq due to the second. Hence the resultant displacement is $Op+Oq$, but this is equal to Or, for $qr=QB=Op$, as is evident from the figure. The same result will be obtained for any other time, so that the resulting motion of the particle will be the projection on the line M_rON_r of the point R, which revolves in a circle of radius OR with the same angular velocity as that of the points in the two circles of reference of the component S.H.Ms. In other words, the resultant motion is a S.H.M. of amplitude OR, the period of which is the same as that of the components. If R is the amplitude of the resultant, then

$$R^2=a_1{}^2+a_2{}^2+2a_1a_2\cos\theta \quad . \quad . \quad . \quad (96)$$

The phase of the resultant is ahead of that of the component of amplitude a_1 by an angle POR which is given by

$$\tan POR=\frac{RC}{OC}=\frac{a_2\sin\theta}{a_1+a_2\cos\theta} \quad . \quad . \quad . \quad (97)$$

86. The Simple Pendulum.—A heavy particle suspended by a perfectly flexible weightless thread forms what is called a simple pendulum. Although it is impossible rigorously to realise a simple pendulum, we may closely approach the required conditions if we suspend a small metal sphere by a long and very thin thread. The distance between the point of support and the centre of the metal sphere will then be the *length* (l) of the simple pendulum.

Let OA (Fig. 147) be a simple pendulum in its position of rest. Suppose the pendulum deflected from its position of rest to the position OB. If m is the mass of the pendulum bob, then we have a force mg acting through B vertically downwards, *i.e.* parallel to OA. This force may be resolved into a component $mg\cos\theta$ along the prolongation of OB, where θ is the angle AOB, and a component $mg\sin\theta$ along BC, the tangent to the circular arc along which the bob moves. The first of these components, namely that along BD, does not tend to bring the pendulum back to its equilibrium position, but simply causes a tension in the supporting thread. The other component, $mg\sin\theta$ along BC, on the other hand, tends to bring the pendulum back to its undisturbed position. Since mg is constant, whatever the displacement, it follows that the force

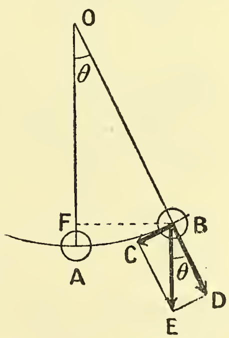


FIG. 147.

of restitution is proportional to the sine of the angle through which the pendulum is displaced. The distance through which the pendulum has been displaced is the length of the arc AB, and is equal to $l\theta$, where l is the length (OB) of the pendulum. The ratio of the force of restitution to the displacement is therefore

$$\frac{mg \sin \theta}{l\theta}$$

In this expression θ is measured in radians (§ 19). Now for an angle of 3°, θ has the value ·05236 radians, and sin θ the value ·05234, while for smaller angles the values of θ and sin θ are even more nearly the same. Hence for small values of the displacement θ we may take $\theta = \sin \theta$, so that the ratio of the force of restitution to the displacement is constant, viz. mg/l. Now in § 84 we saw that when a particle moves under the action of a force such that the force of restitution is proportional to the displacement it will execute a S.H.M. Hence we conclude that the motion of the bob of a simple pendulum will be a S.H.M. so long as the arc over which it swings is small. Further, we saw (equation 92) that the period T of the S.H.M. was given by $T = 2\pi\sqrt{\frac{m}{r}}$ where r is the force of restitution per unit displacement. In the case of the pendulum, when the arc of vibration is small $r = mg/l$. Hence the period is given by

$$T = 2\pi\sqrt{\frac{l}{g}} \quad . \quad . \quad . \quad . \quad (98)$$

It must be remembered that this expression only holds good for a small arc of vibration, say up to two or three degrees; for larger amplitudes the period decreases as the amplitude increases.

At the extremity of its swing the bob of a pendulum is for an instant at rest, so that its kinetic energy is zero. It, however, possesses potential energy, since it is raised up above its lowest position. When the bob is passing through its lowest position, on the other hand, it possesses kinetic but not potential energy. In intermediate positions the bob possesses both potential and kinetic energy. If v is the velocity with which the bob passes through its position of rest A (Fig. 147), its kinetic energy is $\frac{1}{2}mv^2$. The potential energy when at its extreme elongation B is $mg\overline{\text{FA}}$, where FA is the height to which the bob has been raised. But $\overline{\text{FA}} = l - l \cos \theta$, so that the potential energy is $mgl(1 - \cos \theta)$. Since, if we neglect friction, the total energy of the bob must remain the same throughout its motion, we have

$$\tfrac{1}{2}mv^2 = mgl(1 - \cos \theta)$$

or

$$v^2 = 2gl(1 - \cos \theta) \quad . \quad . \quad . \quad (99)$$

from which we can calculate the velocity with which the bob will pass through its position of rest, if after being deflected through an angle θ it is released.

87. The Compound Pendulum. Kater's Pendulum.—In the ideal simple pendulum we have neglected entirely the mass of the string, and have further considered the bob as being so small that during the motion its energy of rotation is negligible compared to its energy of translation. We have now to consider the motion of a pendulum in which these restrictions do not apply, that is, deal with what is called a compound pendulum.

Suppose we have a body of mass M of any shape suspended so that it can rotate about a horizontal axis through S (Fig. 148), and that G is its centre of gravity, which is at a distance h from S. When this pendulum is deflected through an angle θ from its position of rest, the couple tending to *restore* it to its undeflected position is $Mg \times$ GA or $Mgh \sin \theta$. Now in § 19 we saw that if I is the moment of inertia of a body about a given axis, the angular acceleration produced by a couple about this axis is equal to the quotient of the couple by the moment of inertia. Hence the angular acceleration of the pendulum is given by

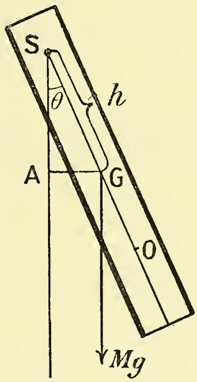


FIG. 148.

$$f' = -\frac{Mgh \sin \theta}{I}$$

or if the deflection is so small that we may put $\sin \theta = \theta$,

$$f' = -\frac{Mgh\,\theta}{I}$$

Since Mgh/I is a constant, the angular acceleration f' is proportional to the angular displacement θ, so long as θ is small, and hence the pendulum will execute an angular S.H.M. If T is the period of the S.H.M., we have from equation (93), p. 217,

$$T = 2\pi\sqrt{-\frac{\theta}{f'}} = 2\pi\sqrt{\frac{I}{Mgh}} \quad . \quad . \quad . \quad (100)$$

If k is the radius of gyration (§ 19) about an axis through G parallel to that about which the body rotates, the moment of inertia about this axis is Mk^2 and that about the axis through S is $Mk^2 + Mh^2$. Hence

$$T = 2\pi\sqrt{\frac{k^2 + h^2}{gh}} \quad . \quad . \quad . \quad (101)$$

Now a simple pendulum of a length l given by $l=\frac{k^2+h^2}{h}$ would have a period $2\pi\sqrt{\frac{k^2+h^2}{gh}}$ equal to that of the compound pendulum. Hence the length, l, of this simple pendulum is called the length of the equivalent simple pendulum. Since $l=h+k^2/h$, the position of the bob of this equivalent simple pendulum would be at a point O on the opposite side of the centre of gravity to the point of suspension S. The point O is called the *centre of oscillation*, and it is the point at which the whole mass of the pendulum could be concentrated without affecting the period. If we call the distance GO h', so that $l=h+h'$, we have

$$h+h'=h+\frac{k^2}{h}$$

or

$$hh'=k^2$$

Next suppose the pendulum to be supported from an axis passing through the point O, the period will now be given by

$$T'=2\pi\sqrt{\frac{k^2+h'^2}{gh'}}$$

or since $h'=k^2/h$,

$$T'=2\pi\sqrt{\frac{k^2+h^2}{gh}}$$

which is equal to T, so that the centre of oscillation, O, is such that the period of the pendulum when vibrating about an axis through this point is the same as that about the parallel axis through the point of suspension S.

Kater's reversible pendulum.

Kater's reversible pendulum, which is used for measuring g, the acceleration of gravity, depends on determining the distance between the centre of suspension and oscillation, by adjusting two knife-edges, one of which is movable, until the period T of the pendulum is the same when the pendulum oscillates about one or the other. If the distance, l, between the knife-edges is then measured, we know that a simple pendulum of length l would have a period T, and hence from the expression

$$T=2\pi\sqrt{\frac{l}{g}}$$

we have

$$g=\frac{4\pi^2 l}{T^2}$$

which gives the value of g.

If a body which is capable of rotation about an axis through S (Fig. 148) is struck a blow in a direction passing through the centre of oscillation, the body will be set into rotation about the axis but no effect of the blow will be felt at the axis. In other words, if a body be perfectly free to rotate about any axis and it is struck a blow at the centre of oscillation O, it will start to rotate about an axis through the point S. For this reason the point O is called the centre of percussion with reference to an axis through S. A familiar example of the above property of the centre of oscillation occurs when a ball is struck with a cricket-bat. If the ball is struck with a certain part of the bat, the impact is not felt by the hands of the striker; if, however, the bat is struck either higher up or lower down, a distinct "sting" is felt in the hands. When no sting is felt, the ball has been struck with the part of the bat which is the centre of oscillation corresponding to the centre of suspension where the bat is held in the hands.

Torsion pendulum.

A particular form of compound pendulum with which we shall have to do later is formed by a body, say a cylindrical rod of metal, suspended by a wire so that the axis of the rod is horizontal. If the top of the wire is held fixed and the lower end is twisted through an angle θ, a couple, due to the torsion of the wire, will be produced which will tend to bring the body back to its position of equilibrium. If the couple produced when the body is twisted through one radian is c, the couple when the twist is θ will be $c\theta$, and hence since the couple is proportional to the displacement the vibrations will be simple harmonic. And if I is the moment of inertia of the rod about the wire as an axis, we have the angular acceleration is

$$-\frac{c\theta}{I}$$

and hence from equation (93), p. 217,

$$T=2\pi\sqrt{\frac{I}{c}}$$

Such an arrangement is called a torsion pendulum, and it is important to observe that since the couple produced by even a large rotation θ is proportional to θ so long as the wire is not short, the period is the same for both small and large amplitudes.

In the case of the ordinary pendulum it will be remembered that the force of restitution varied as $\sin\theta$, so that the period is only independent of the amplitude so long as θ is so small that we can put $\sin\theta=\theta$.

As a consequence of the above, if a clock is to keep good time the arc over which the pendulum swings must either be kept very small, or if a fairly large arc is employed, this arc must always be the same, whether the clock has just been wound up or not. In a watch or chronometer,

however, we have to do with a modified form of torsion pendulum, for the couple produced by winding up the hair-spring is very nearly proportional to the twist. Hence the period does not depend to a great extent on the amplitude of the vibrations of the balance-wheel.

Compensated pendulum.

Since the period of a pendulum depends on the length, it is important that the length of the pendulum of a clock should remain always the same. Now when the temperature rises most bodies expand, so that on this account the length of the pendulum will increase and hence the clock will lose time. To get over this effect pendulums have been designed composed of two metals having different coefficients of expansion. One form of such pendulums, called a gridiron pendulum, is shown diagrammatically in Fig. 149. The rods AB and CD are made of steel, and the rods BC of brass. Suppose L_s is the sum of the lengths of one of the rods AB and of CD, and a_s is the coefficient of linear expansion (§ 60) of steel, while L_b is the length of one of the brass rods BC, the coefficient of expansion being a_b. When the temperature rises by $t°$, the steel will expand by an amount $L_s a_s t$, and hence on this account the bob will be lowered by this amount. The brass rods will expand by an amount $L_b a_b t$, and the bob will on this account be raised by this amount. Hence the pendulum will be unaltered in length if

FIG. 149.

$$L_s a_s t = L_b a_b t,$$

or

$$\frac{L_s}{L_b} = \frac{a_b}{a_s}.$$

Since the coefficient of expansion of brass is about 1·7 times that of steel, the total length of the steel rods ought to be 1·7 times the total length of the brass rods. By using as the two metals a nickel steel called invar, which has a coefficient of expansion of about $0{\cdot}0087 \times 10^{-4}$, and zinc which has a coefficient of $0{\cdot}292 \times 10^{-4}$, so that the ratio of expansion is 33·5, an invar pendulum rod of 39 inches length requires a zinc compensator of only a little over an inch in length. The pendulum is thus much simplified, for the bob slides on the invar rod and rests on the top of a zinc washer about an inch thick, this washer resting on a cross-bar attached to the bottom of the invar rod.

88. Forced and Free Vibrations. Resonance.—In the preceding sections we have considered the motion of a pendulum of which the bob after being pulled aside is allowed to vibrate unacted upon by any force other than gravity. That is, we have studied what we called the *natural* or *free* vibrations of the pendulum. We have now to consider what will be the nature of the motion if a periodic external force is applied to a pendulum.

Suppose that a pendulum with a very heavy bob A (Fig. 150) is suspended from an axis B so that it can vibrate in a plane perpendicular to the paper, and that from a rod C, attached near the axis of suspension, is suspended a small simple pendulum D. In this way the point of support of the simple pendulum will be given a small to-and-fro motion, and hence the simple pendulum will be acted upon by a periodic force, the period being that of the large pendulum. Since the bob A is very large compared to D, the small pendulum will not appreciably affect the motion of the large pendulum. Let T be the period of the large pendulum and L the length of the equivalent simple pendulum. If the length l of the simple pendulum CD is less than L, then when the big pendulum is started the small pendulum will be set in motion, and the amplitude of the motion will gradually increase, then decrease, and then increase again. These alternate increases and decreases will gradually get less marked, till finally the amplitude will remain constant. It will then be found that the period of the small pendulum is exactly the same as that of the large one, and that the phases of the two are very nearly the same, that is, when the bob A is moving to the right so is bob D. Now since the length of the pendulum CD is less than that of the equivalent simple pendulum corresponding to A, it follows that the natural period of CD is less than that of A. Hence, owing to the influence of the periodic force, the pendulum CD has been caused to vibrate in a period which is not its natural period. The vibrations produced in this way are called *forced vibrations*, and the characteristic is that they have a period *equal to that of the applied periodic force.*

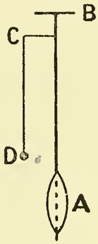


FIG. 150.

Forced vibrations of a simple pendulum.

If the length of CD is now made greater than L, so that the natural period is longer than the period of the applied force, then at the start the same alternations in the amplitude will occur; but when these have died out, and the amplitude has become constant, the period of CD will again be the same as that of A. In this case, however, the phase of CD will differ from that of A by nearly 180°, that is, when A is moving to the right, D will be moving to the left.

Finally, if the length of CD is exactly equal to L, there will be no alternations of amplitude when the motion starts, but the amplitude will steadily increase and finally reach a value very much greater than that produced in the two preceding cases. The periods of the two pendulums will still be the same, but in this case the natural period of CD agrees with that of the force. Thus when a periodic force acts on a pendulum, or any other body capable of executing vibrations, and the period of the force agrees with the natural period, the amplitude of the resulting

motion is much larger than that which occurs when the period of the force differs from the natural period, and we are said to be dealing with a case of *resonance.*

Resonance.

This phenomenon of resonance is of very great practical importance, particularly to engineers, for on its account a quite small periodic force, if it happens to have the same period as the natural period of some structure or machine, may set up vibrations of very great magnitude. Now these vibrations are generally accompanied by stresses set up in the structure, in fact it is the stresses which supply the force of restitution which is essential for the production of natural vibrations, and hence if the amplitude becomes great the stresses may be beyond the elastic limit, and hence the structure be damaged. Thus it is important to take care that the period of the applied force never agrees with the natural period of the structure.

The effect of resonance is often very marked in ships fitted with reciprocating engines. Owing to the inertia of the reciprocating masses, such as pistons, &c., a periodic force is applied to the hull, and if the period agrees with a natural period of the hull, marked vibration is set up. If, however, the engines are made to rotate at a slightly lower or higher speed, so that resonance no longer occurs, the vibration set up is very much reduced.

CHAPTER II

WAVE-MOTION

89. Wave-Motion.—We have in the preceding chapter considered the periodic motion of a single particle or of a body, such as a pendulum, where the phase of the motion of all the particles which build up the body is the same; we have now to consider in some detail the resultant motion when the various particles of a medium are executing periodic motions, but the phases of the motions of the various particles are not the same for all, but are related to one another in certain definite ways.

Suppose we have a number of particles arranged, when at rest, at equal distances along a line AB (Fig. 151), and that these particles all

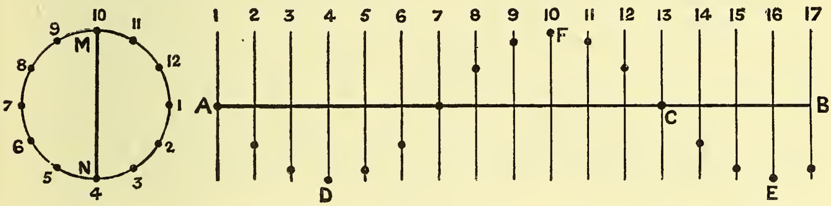


FIG. 151.

execute S.H.Ms. of equal amplitude and period along lines at right angles to AB, but in such a way that the phase of each successive particle, counting from A, differs from that of the preceding particle by a constant amount.

Thus if the constant difference in phase is 30°, when the particle 1 is at its median position, the position of the others will be as shown by the dots in the figure. The displacement of particle 2 at any moment is equal to the displacement of particle 1 at 1/12 of the periodic time (T) later, since 30° is 1/12 of 360°. Similarly, particle 3 is displaced to the amount that particle 1 will be at $2T/12$ from the start, and so on. Hence the curve drawn through the positions of the particles at any instant will be a harmonic curve (§ 84). Particle 13 will at every moment be in exactly the same state as particle 1, particle 14 as

particle 2, and so on; for, as their phases differ by a whole period, they will be equally displaced and *moving in the same direction.*

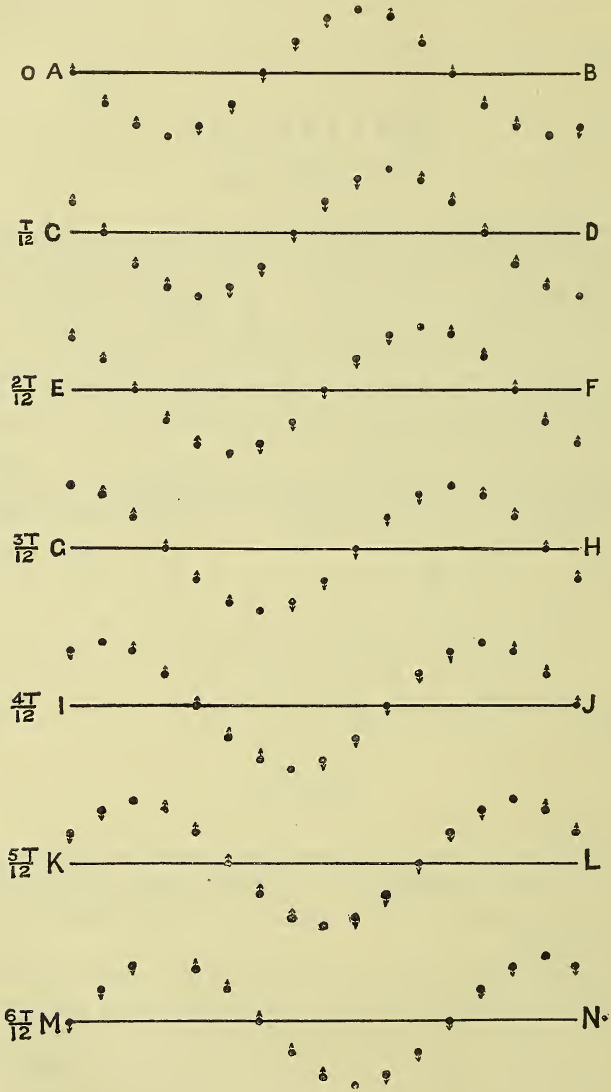


FIG. 152.

In Fig. 152 the positions of the particles are shown at successive intervals of $\frac{T}{12}$ up to half a complete vibration from the positions de-

picted in the first line, the direction of motion at the given instant being indicated by an arrow-head. It will be seen that the curve drawn through the particles can in each case be obtained by displacing the curve for the preceding configuration to the right, and hence, as the motion goes on, the curve connecting the particles appears to move steadily to the right. The distance through which it moves during one complete period of one of the moving particles is equal to the distance between two particles which are moving at every instant in the same direction and are equally displaced on the same side of their mean positions. This distance through which the curve, called in this case a wave, moves during a complete period of one of the moving particles is called the *wave-length* of the motion. The wave-length may also be defined as the distance between one particle and the next one that is displaced from its mean position to the same extent and is moving in the same direction, that is, between two consecutive particles which are in the same phase. Thus in Fig. 151 the wave-length is equal to AC or DE.

Transverse wave.

Although the form of the wave is similar to the harmonic curve, it must be remembered that the harmonic curve represents the successive displacements of a single particle, the abscissæ representing time, while the wave-form curve represents the simultaneous positions of a number of particles, the abscissæ being the distance of the mean positions of the particles measured from some fixed point. However, as all the particles move in exactly the same way, and in one whole wave-length we shall have an example of a particle in every phase of this motion, we may look upon the wave-curve as also showing us what the displacement of each particle will be at different times.

A point on the wave such as F (Fig. 151), at which the particle is at its maximum positive displacement, is called a *crest*, while a point such as D or E, where the displacement has its maximum negative value, is called a *trough*. The positions of the crests and troughs appear to travel towards the right as the motion of the particles continues.

This translatory motion of the wave is not accompanied by the translation of the particles themselves, that is, although each particle moves to and fro along its own little path, yet its mean position during a complete oscillation remains unaltered. We may, therefore, define a wave as a form of disturbance which travels through a medium, and is due to the parts of the medium performing in succession certain periodic motions about their mean positions.

In the case of wave-motion considered above, the particles all vibrate at right angles to the direction in which the wave moves, and this form of wave-motion is said to be due to *transverse vibrations*. If the motion

of each particle takes place in the direction in which the wave moves, then the vibration is said to be *longitudinal.*

Longitudinal wave.

At AB (Fig. 153) the undisplaced positions of the particles are shown. If each particle now executes a S.H.M. in the direction AB, the period and amplitude being the same for all, but the phase of each particle being 30° behind that of the preceding particle, then when particle 1 is passing through its mean position and moving towards the right, CD will represent the positions of the other particles. The positions of the particles are also shown at successive intervals of 1/12 of the period of the S.H.M. of each particle for half a complete period.

In this form of wave-motion the distances between adjacent particles alter, so that the particles are alternately crowded together and spread

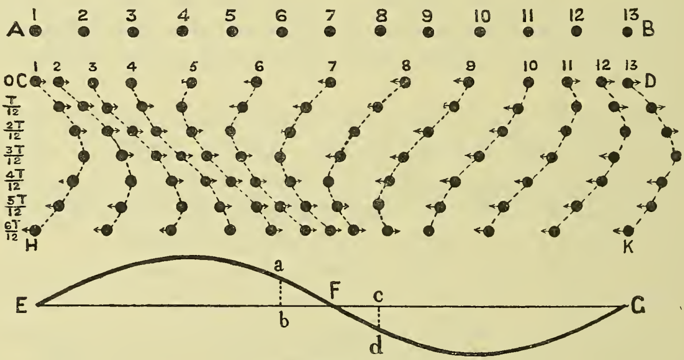


FIG. 153.

out. A point where at any instant the crowding together is a maximum is called a condensation, while a point where the distance between adjacent particles is a maximum is called a rarefaction. These play the same parts in longitudinal wave-motion as do the crests and troughs in transverse wave-motion.

The definition of wave-length, given with reference to transverse vibrations, applies also to longitudinal vibrations. The most convenient manner of studying longitudinal vibrations is to employ a curve of which the ordinates indicate the displacements from their mean positions of the different particles at any time. Such a curve is obtained if, at the mean or undisturbed position of each particle, we erect a perpendicular in the positive or negative direction according as the displacement of the particle is in the positive or negative direction, and having a height equal to the displacement of the particle from its mean position. The curve obtained by joining the extremities of these ordinates is shown at EFG

(Fig. 153), the corresponding positions of the disturbed and undisturbed particles being shown at HK and AB. This curve is a harmonic curve, and the points where the curve cuts the axis correspond to the places where the particles are most crowded together, or most spread out. For at F the particle 7 is at its mean position, while the particle 6, since the corresponding ordinate *ab* of the curve is positive, is displaced to the right by an amount equal to this ordinate, and the particle 8 is to the left of its median position by an amount equal to the ordinate *cd* of the curve; so that the particles are here crowded together. In the same way the particles at E and G are separated to a maximum extent. Hence F corresponds to a condensation, while E and G correspond to rarefactions. The distance between two adjacent rarefactions, such as E and G, or between two condensations, is equal to the wave-length of the wave-motion, while the distance between a rarefaction and the adjacent condensation is equal to half a wave-length.

The speed at which the crest or trough in the case of a transverse wave, or the condensation or rarefaction in a longitudinal wave, moves through the medium is called the velocity of propagation of the wave-motion.

While particle 4 in Figs. 151 and 152 is making a complete oscillation, the trough of the wave will travel to the right to particle 16, that is, through a distance equal to the wave-length l. In the same way, while particle 1 (Fig. 153) is making a complete oscillation, the condensation will travel from C to D, that is, through a distance equal to the wave-length.

Hence if T is the time each particle takes to complete one oscillation, in this time the wave will move through a distance equal to the wave-length. Thus if v is the velocity of propagation of the wave, we have

$$v=l/T \quad . \quad . \quad . \quad . \quad . \quad (102)$$

Each time that particle 10 (Fig. 151) reaches its maximum positive elongation, a crest will be passing at F, so that the interval between the passage of two successive crests is T. Thus if n is the number of crests which pass F in a second, we have $n=1/T$. The same remark applies to any other particle, whether the motion is transverse or longitudinal, and the quantity n is called the frequency of the waves. Thus

$$v=nl \quad . \quad . \quad . \quad . \quad . \quad (103)$$

In the case of a wave in a medium where the force of restitution called into play when a portion of the medium is displaced is due to the elasticity of the medium, it can be shown that the velocity of a

wave is given by the following expression, which was first obtained by Newton—

$$v=\sqrt{\frac{E}{D}} \quad . \quad . \quad . \quad . \quad (104)$$

where E is the elasticity of the medium and D is its density.

The velocity with which a group of waves moves into an undisturbed portion of the medium is not necessarily equal to the velocity of the individual waves. Thus in the case of gravitational waves on a liquid, the individual waves travel twice as fast as does the front of the disturbance. Thus if we watch a short train of waves moving into still water, the waves will appear to move through the group, dying out in front, and fresh waves appearing in the rear of the group. It can be shown that whenever the velocity of the waves varies with the wave-length, the group velocity is different from the wave velocity.

The study of waves being of very great importance in physics—for, as we shall see, sound, light, radiant heat, and many electro-magnetic phenomena are propagated by wave-motions—it will be advisable to spend some time considering this form of motion. It will add to the interest, and also to the clearness, of the study of a wave-motion if we illustrate the various points by reference to some particular form of wave-motion. Now the waves which constitute sound, light, and heat are invisible, and so for the purposes of illustration it will be better to consider the waves which may be produced at the surface of a liquid, for such waves may, with suitable arrangements, be seen by the eye.

90. Waves on the Surface of a Liquid.—In order that a wave may be produced it is necessary that the successive particles which constitute the medium in which the wave is propagated should each in succession go through a periodic motion. Now when considering the motion of a pendulum, we showed that the reason it executes its periodic motion is that, when the bob is displaced, a force acts on the bob tending to bring it back to its position of rest. Hence when dealing with the production of a wave-motion in a medium, we must consider how the force of restitution on the particles of the medium, which is necessary for the production of the periodic motion of these particles, is brought about.

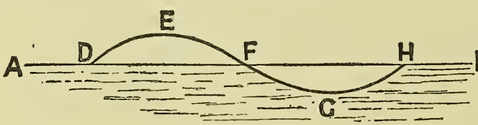


Fig. 154.

Let AB (Fig. 154) represent the plane surface of a liquid at rest. Now suppose by some means we cause the liquid to be heaped up in the form DEF, or scooped out into a hollow FGH, so that the liquid particles are displaced from their positions of rest. Then, owing to the action of

gravity, the particles in the portion DEF of the liquid will move back towards the level surface AB, while the particles which have been forced down owing to the production of the hollow FGH will move up. Thus when the particles of a liquid are moved, so that a portion of the surface is displaced either above or below the level of the general surface, owing to gravity a force will act tending to bring the surface back to its undisturbed position. We have therefore the conditions suitable for the production of waves on the surface of the liquid, and the existence of these waves being due to the action of gravity, they are called gravitational waves. The large waves seen on the surface of the sea are well-known examples of gravitational waves. Gravity, as was, however, first pointed out by Lord Kelvin, is not the only cause tending to bring the surface of the liquid back to its undisturbed position. There is a second cause acting, namely, the surface tension (§ 52) of the surface film of the liquid. This surface tension acts as if there were a thin elastic membrane stretched over the surface, and it is evident that the effect of such a stretched membrane will be to tend to flatten down the portion DEF of the disturbed surface of the liquid, and to level up the portion FGH, so that in the surface tension we have also a force of restitution acting on the displaced liquid particles. Now the pressure due to the surface tension increases with increase of the curvature of the surface, so that, since the magnitude of the surface tension is small, it is only when we are dealing with waves in which the curvature is very great that we need take account of the surface tension.

It can be shown, although to do so would lead us beyond the scope of this book, that if v is the velocity of a wave along the surface of a liquid of which the depth is not less than the wave-length l, and D and T are the density and surface tension of the liquid, then

$$v^2 = \frac{gl}{2\pi} + \frac{2\pi T}{lD} \qquad . \quad . \quad . \quad . \quad (105)$$

From this expression it follows at once that if the wave-length l is great, the fraction $2\pi T/lD$ is small compared to $gl/2\pi$, and hence may be neglected. The fact that l is great shows that the curvature of the surface must be small, so that this result is what we should expect. On the other hand, if l is small, then $2\pi T/lD$ is great compared to $gl/2\pi$, so that in this case surface tension plays the important part in the propagation of the waves. Such waves, in which the greater part of the force of restitution is due to surface tension, are called capillary waves or ripples.

For waves of wave-length greater than about 4 inches or 10 cm. the term $2\pi T/lD$ may be neglected, while for waves of wave-length less than 0·1 inch or 3 mm. the term $gl/2\pi$ may be neglected. For waves having

wave-lengths between these two limits, we have to take into account both the effect of gravity and of surface tension.

Since the velocity due to gravity alone increases as l *increases*, and that due to surface tension alone increases as l *decreases*, it follows that there must be a certain wave-length for which the velocity is a minimum. For water the minimum velocity is 23 cm. per second, or 9 inches per second.

In the case of waves for which the wave-length l is so great that we may neglect the effect of surface tension, we have

$$v^2 = \frac{gl}{2\pi} \quad . \quad . \quad . \quad . \quad (106)$$

It will be observed that the density of the liquid is not involved in the expression for the velocity. The reason for this is the same as that which explains why it is that the period of a pendulum is independent of the mass of the bob, namely, that although the mass of the liquid to be moved is proportional to the density, yet, since the force of restitution is also proportional to the density, for it is the weight of the raised portion of the liquid, the ratio of the force of restitution to the mass to be moved is the same for all liquids, and therefore the velocity of the waves is the same.

If the depth of the liquid is considerably less than the wave-length, the velocity is less than that given above, and is given by

$$v^2 = gd \quad . \quad . \quad . \quad . \quad (107)$$

where d is the depth of the liquid. One effect of the decreased velocity in shallow water is to make the waves in the neighbourhood of a shelving beach always move in a direction perpendicular to the shore, although at some distance out to sea they may be moving in quite a different direction. The reason is that when a wave which is moving in a direction inclined to the shore-line reaches shallow water, the end of the wave which first reaches the shallow moves more slowly than the parts which are still moving in deep water. Thus the wave gradually wheels round till it becomes nearly parallel to the shore.

In the case of a wave in deep water, the individual particles of the water describe circles in vertical planes as illustrated in Fig. 155, where the form of the wave is shown at two instants corresponding to an interval of one-twelfth of the periodic time. Thus when a particle is on the crest, A, of the wave, it is moving in the direction in which the wave is moving, while when it is in the trough, B, it is moving in the opposite direction to that in which the wave is moving. As we go down from the surface the particles still move in circles, but the radii grow

smaller and smaller, till at a depth equal to the wave-length the radius of the circles is only about 1/500th of what it is at the surface. In shallow waters the paths of the individual particles are ellipses with their major axes horizontal. In this case the horizontal axes of the ellipses are approximately the same for particles at all depths. The vertical axes, however, decrease with the depth, till at the bottom they vanish, and the particles move backwards and forwards along straight lines.

When the height of the crest of a wave above the undisturbed level of the water is equal to the depth of the undisturbed water at the point, the particles at the crest will be moving forward with the same velocity as the wave, and the wave will be unstable and "break." As a wave comes into shallow water the wave-length decreases, for the velocity decreases as the depth of water decreases, and the frequency (n) must remain the same, that is, the number of waves which pass a given point in one second, and $v = nl$. The effect of this shortening of the wave-length is to make the amplitude of the waves greater. This goes on, till finally the unstable condition is reached, and the wave breaks.

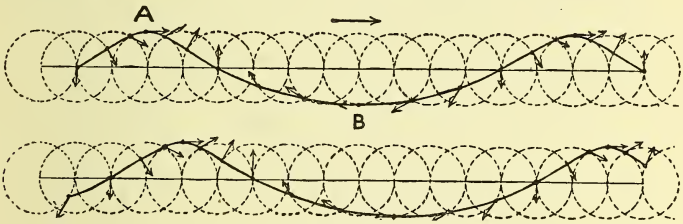


FIG. 155.

In the case of waves of which the wave-length is less than 4 mm., we have

$$v^2 = \frac{2\pi T}{lD} \quad . \quad . \quad . \quad . \quad (108)$$

Hence both T and D depend on the nature of the liquid, so that the velocity of capillary waves is different in different liquids. If n is the frequency of the waves,

$$v = nl,$$

and

$$n^2 l^2 = 2\pi T / lD,$$

or

$$T = n^2 l^3 D / 2\pi \quad . \quad . \quad . \quad . \quad (109)$$

Thus if the frequency n is known, and we measure the wave-length, we can calculate the surface tension T. Lord Rayleigh has used this method for measuring the surface tension. The waves were produced by a fine style attached to the prong of a tuning-fork which dipped into

the liquid. Thus the frequency of the waves was equal to the frequency of the fork, and was therefore known.

91. Interference of Waves. Huyghens's Construction.— If we have two systems of waves passing over the same surface of water, each will produce the same effect as if it were alone present, so that the actual displacement of any surface particle of the water at a given instant is the algebraical sum of the displacements it would have, at that instant, due to each set of waves separately. The resultant motion is thus obtained by compounding the two separate wave-motions, just as in § 85 we obtained the resultant motion of a point when moving with two simple harmonic motions.

This separate existence of two sets of waves is one of everyday observation, when two stones are thrown into still water. Each stone

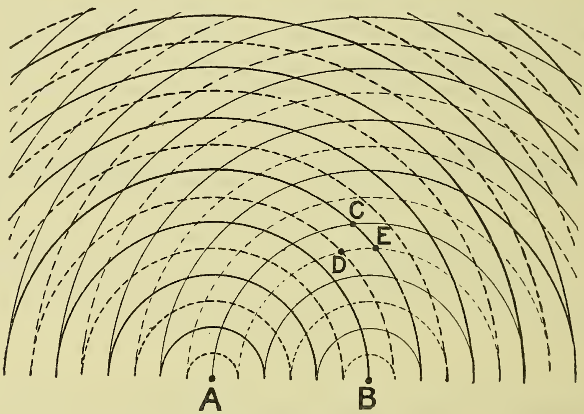


Fig. 156.

will produce a set of waves which travel out in ever-widening circles, and the circular waves due to one stone will pass unchanged through the waves due to the other.

Suppose we have a style attached to one of the prongs of a tuning-fork dipping into a vessel containing a liquid, say mercury. When the fork is in motion the style will produce a system of waves which will move out in circles from the point where the style enters the mercury. The radius of each of the circular waves will increase at the rate given by

$$v = nl,$$

where n is the frequency of the fork, and l the wave-length as given by equation 109—

$$l^3 = 2\pi T/n^2D.$$

Let the position of the waves at a given instant be as represented by the circles, with A as centre, in Fig. 156, where the heavy full lines represent the crests, and the heavy dotted lines the hollows. The waves in only half the circumference are drawn, in order to save space. Now, suppose there is a second style attached to the same prong touching the mercury surface at B, so that whenever a crest starts from A an equal crest will start from B, and so on. The position of the waves due to B alone are shown in the figure by the light lines. The wave-length and velocity of the waves starting from B will be the same as those starting from A. In order to obtain the actual condition of the surface due to the combined action of the two sets of waves, we have at every point to add together the displacements due to each set. Thus at the point C we have a crest of each set of waves, so that the upward displacement

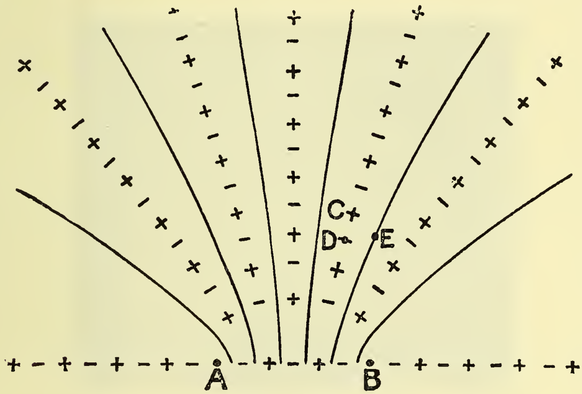


FIG. 157.

here is twice the maximum displacement due to either set separately. At D, in the same way, we have a trough due to each set, and the downward displacement is double that due to either set of waves alone. At E, however, we have a crest due to the waves starting from A, and a trough due to the waves starting from B. The result is that the particle at E is undisplaced, for the upward displacement due to one set of waves is just neutralised by the downward displacement due to the other. In Fig. 157 lines are drawn through the points which are undisturbed, while a + sign marks the points where there is maximum upward displacement, and a − sign the points where there is a maximum downward displacement.

Next, suppose we again draw the figure for a time equal to half a period later. Each of the waves will have travelled out through a distance equal to half a wave-length, and so a crest will now occupy the

position previously occupied by a trough, and *vice versâ*. Thus the conditions are still represented by Fig. 156, where, however, the dotted lines now represent the crests and the full lines the troughs. The point C is now at a trough, and the point D at the crest of the disturbance due to both sets of waves. The point E is, however, still at rest, for now it is at the trough of a wave from A, and at a crest of a wave from B. It will also be found that all parts of the liquid surface along the lines drawn in Fig. 157 are still at rest. In this way it can be shown that, owing to the joint action of the two sets of waves, we have certain portions of the mercury surface which are permanently at rest, although, if either set of waves acted alone, these parts would be disturbed by the passage of waves. This phenomenon, of a state of rest being produced by the combined action of two sets of waves,

FIG. 158.

is called *interference*, and we shall find that it plays a very prominent part in many natural phenomena.

That we actually do get these lines of no disturbance in the case of capillary waves can be seen by eye, for, although the individual waves cannot be distinguished on account of their rapid motion, yet the undisturbed portions of the surface appear brighter than the rest, and we see a pattern similar to Fig. 158. If, instead of looking at the surface, we take an instantaneous photograph, then, as shown in Fig. 159, we see not only the lines of no disturbance but also the waves between these lines, which are due to the combination of the two sets of waves.

In the place of two centres of disturbance, such as those considered above, suppose we have a number placed in a line AB (Fig. 160), where,

for simplicity, the position of the crests only are shown for three waves due to each centre. It will be observed that along the lines which touch the circular waves due to all the centres of disturbance, the waves all work together to form a crest, while at other places we have a crest due to one set coinciding with a trough due to others. The result is that, except along the tangents, there will be very little disturbance. Of course half-way between the crests we should, if we drew the circles to represent the positions of the troughs, have a resultant trough which would also be a straight line parallel to AB. It will be understood how, if the number of centres of disturbance between A and B was made very great, the resultant effect would be a series of waves which would form straight lines parallel to AB, and would move out over the surface of the liquid in a direction at right angles

Huyghens's construction.

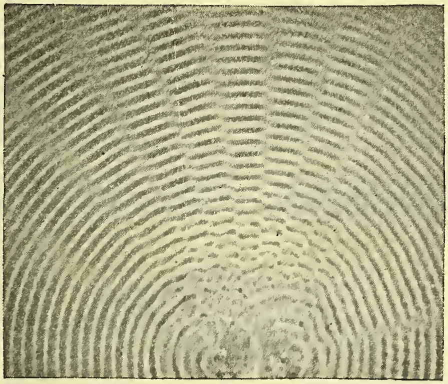


FIG. 159.

to AB. We have thus arrived at an explanation of the formation of a plane wave, that is, one in which the crests and troughs are straight lines, and may look upon it as produced by the action of an infinite number of disturbing centres placed along the straight line AB. Such a plane wave is obtained in practice if, in place of a style, we attach a flat glass plate to the prong of a tuning-fork, so that the edge dips into the mercury. The construction for finding the position of one of the crests, that is, of the wave-front,[1] at a time t by means of the tangent to a series of circles, the radius of each of which is equal to the space passed over by the wave in a time t, is due to Huyghens, and is known

[1] A crest or a trough, or in fact any line or surface such that all parts of the medium on this line or surface are at every instant moving with the same phase, is called a wave-front.

as Huyghens's construction for the wave-front. The centres of disturbance need not lie on a straight line. Thus suppose they lie on a circle ABC (Fig. 161), of which the centre is D and radius R. If now from all points along the circumference of this circle we describe circles of radius r,

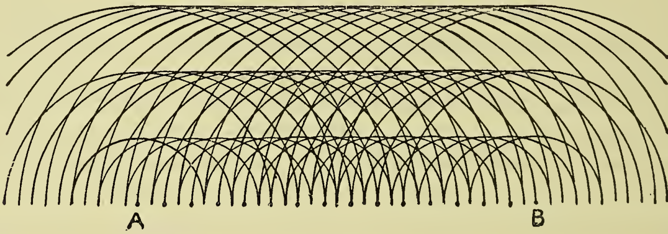


FIG. 160.

where r is the distance the wave will travel through in a time t, we shall get the wave-front at a time t by drawing a line touching all these circles. This tangent is evidently a circle of radius $R+r$, having its centre at D. Now if a disturbance were produced at D, we know that in a time R/v, where v is the wave velocity, the wave-front would be the circle ABC, while at a time t later it would be the circle EFG, so that if we

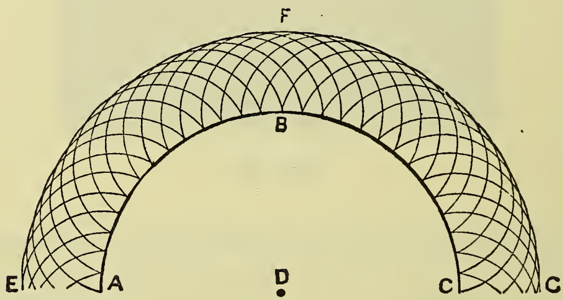


FIG. 161.

like we may look on each point on the wave-front ABC as being a centre of disturbance, and consider that it is due to the combined action of these secondary centres of disturbance that the wave-front EFG is produced, and not directly by the disturbance at D. We shall find Huyghens's construction of considerable use in future for finding the position of the wave-front at any time subsequent to that at which the wave-front has some given position.

It will be perceived from the above construction that the direction in

which any given portion of the wave is moving is along the normal to the wave-front at that point. So long as the medium in which the wave is propagated is isotropic, this result is quite general. If the elasticity of the medium is different in different directions, the direction of wave propagation is not necessarily along the wave-normal. Except when considering double refraction we shall be solely concerned with waves in isotropic media, and hence shall always be dealing with waves where the direction of propagation is along the wave-normal. A line which is everywhere normal to the wave-front is called a *ray*, and it everywhere indicates the direction in which the wave is travelling.

In isotropic media direction of propagation is along the wave-normal.

92. Reflection of Waves.—When a wave meets an obstacle it is not in general annihilated, but its direction of propagation becomes

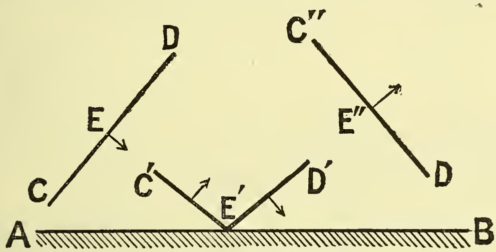


FIG. 162.

altered. This phenomenon is called *reflection*. Thus suppose AB (Fig. 162) represents a wall limiting a stretch of water, and CED represents a wave moving in the direction of the arrow. As this wave moves forward, first the end C and then each part in succession will meet the wall and will be reflected, so that, in addition to the portion E′D′ of the wave moving in the original direction, there will be a reflected portion C′E′ moving in the direction shown by the arrow; while, finally, when the whole of the wave has been reflected we shall have a wave C″E″D″, which will move off in the direction indicated by the arrow. In order to find the connection between the inclination of the incident wave-front to the reflecting surface and that of the reflected wave-front, consider the incident wave CL (Fig. 163), of which the end C has just reached the reflecting surface AB. If there had been no reflecting surface, then, when the other end of the wave L reached G, the wave would have occupied the position HG. Let us start reckoning time from the instant when the wave reaches C, that is, from the instant when the wave is in the position CL, and suppose that the time taken by the point L on the wave to reach G is t, so that if v is the velocity with which the wave moves,

$$t = \frac{LG}{v}$$

Now each of the lines CH, DD‴, EE‴, FF‴, is equal to LG, and represent the distance moved through by the wave in the time t. When the end C of the wave reaches the reflecting wall, we may consider that C becomes a centre of disturbance, so that in a time t the wave produced by this centre will form a circle of which the radius r is given by $r = vt$, that is, r = LG = CH. Hence, if a circle is described with C as centre and radius CH, it will represent the position of the reflected wave from the point C at the time t. Next, consider some other point D on the wave. It will have to move over the distance DD′ before it meets the reflector, and when it does so, we may regard it as becoming a centre of disturbance. Since, if it were not for the reflection, in the time t, the point D on the wave would travel to D‴, it is clear that in the time t the wave will travel up

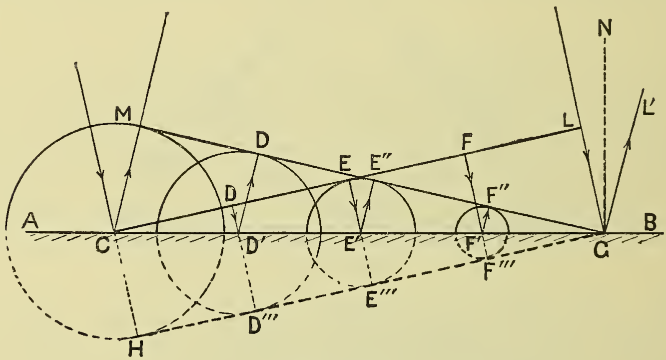


Fig. 163.

to D′, *i.e.* through the distance DD′, and then have time to travel through a further distance equal to D′D‴. Therefore, if a circle is described round D′ as centre, with D′D‴ as radius, this circle will represent the position, at the time t, of the reflected disturbance produced by the point D on the wave.

Proceeding in this way, we can show that the reflected waves formed by each point of the incident wave meeting the reflecting surface touch a straight line MG, for all the centres of the circular waves lie on the straight line AB, and the circles which show the position of the reflected waves at the time t all touch the straight line HG. Hence by Huyghens's principle the line MG represents the position of the reflected wave-front at the time t after the instant when it occupied the position CL.

The two triangles CLG, GCM are equal, for LG = CM, and CG is common, and the angles CLG, GMC are each right angles. Hence the angle LCG, made by the incident wave-front with the reflecting surface AB, is equal

to the angle MGC, made by the reflected wave-front with the reflecting surface.

The line LG shows the direction in which the incident wave moves, and is the incident ray, while the line GL′ shows the direction in which the reflected wave moves, and is the reflected ray, for these lines are perpendicular to their respective wave-fronts. Since the wave-fronts make equal angles with AB, these perpendiculars must also make equal angles with AB, so that the angle LGC is equal to the angle L′GB. Hence, if GN is the normal to AB at G, that is, a line drawn through G perpendicular to the reflecting surface, the angle LGN must be equal to the angle L′GN. Now the angle LGN, between the direction of motion of the incident wave and the normal to the reflecting surface, is called the angle of incidence, while the angle L′GN, between the direction of motion of the reflected wave and the normal, is called the angle of reflection. Hence we see that when a wave is reflected, the angle of incidence is equal to the angle of reflection.

We have hitherto considered the reflection of a single wave, so that at no time did the incident and reflected waves affect the same particles simultaneously. If, however, we are dealing not with a single wave but with a train of waves, then a given point in the liquid may be affected by one of the incident waves, and at the same time by one of the preceding waves which is returning after reflection. The result of this simultaneous action of two sets of waves, the incident and the reflected, is that interference may take place. We shall only consider the simplest case, namely, that where the direction of the incident waves, and therefore also that of the reflected waves, is at right angles to the reflecting surface.

Stationary waves.

Let AB (Fig. 164) represent a vertical section of the reflecting surface, and let the waved line CD represent a section through the incident waves. Each of these waves, as it reaches the obstacle, will be reflected, so that we shall have a train of reflected waves travelling away from AB, which for clearness are shown separate at EF.

Since it is evident that when a crest of the incident waves reaches the obstacle a crest will be produced on the reflected wave, and as at the instant for which the figure is drawn a crest on the incident wave is at C, we must have a crest on the reflected wave at E. Owing to the combined action of the incident and reflected waves, the form taken by the water surface is shown at GH, in which the displacement at any point is the sum of the displacements due to the two systems of waves separately.

If T is the period of the waves, then in a time $T/4$ the incident waves will have moved through a distance equal to a quarter of the wave-length, l, to the left, for the wave travels over a space equal to the

wave-length during the period; also the reflected waves will have travelled through a distance $l/4$ to the right, as shown at C′D′ and E′F′. Under the combined action of the two sets of waves the whole surface will be momentarily in its position of rest, as shown at G′H′, for it will be noticed that the displacement due to the reflected waves is just suffi-

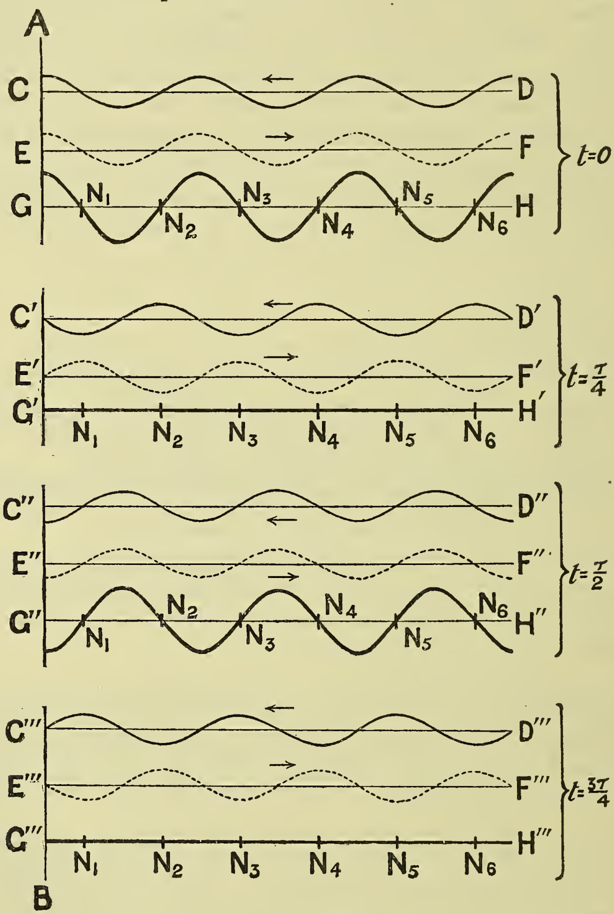


FIG. 164.

cient to neutralise the displacement due to the incident waves. Similarly the actual state of the water-surface, at times $T/2$ and $3T/4$, is shown at G″H″ and G‴H‴. If these curves are examined, it will be seen that there are certain points, N_1, N_2, N_3, &c., on the surface of the liquid which remain permanently at rest, owing to the interference between the incident and reflected waves. These points are called *nodes*. Half-

way between each node the water-surface swings up and down to a maximum extent, and these points are called *loops* or *antinodes.* The portions of the surface between the nodes and loops move up and down, the amplitude of the movement gradually decreasing from the loop to the node. Thus at one instant we have a series of alternate crests and troughs at the loops, then the surface flattens out, and immediately after a series of troughs and crests appear at the loops, and so on; and the character of the disturbance is quite different from ordinary waves, for there is no progressive movement of the crests and troughs. These waves which retain their position unaltered are called *stationary waves*, and, as we shall find later, they play an important part in many phenomena which involve wave-motion.

It is immediately evident, from Fig. 164, that the distance between consecutive nodes is equal to half the wave-length. Hence if we measure the distance between the nodes, and know the frequency of the waves, we can calculate the velocity with which they travel.

The nodes are points at which the disturbance due to the reflected waves is *always* equal and *opposite* to that due to the incident waves. The loops, on the other hand, are points where the disturbance due to the reflected is equal to, and in the *same* direction as, that due to the direct waves. A consideration of Fig. 164 will show that the portions of the medium on opposite sides of a node are always moving in opposite directions, or are displaced on opposite sides of their positions of rest.

The stationary waves considered above are produced owing to interference; there are, however, another class of waves which *appear* at rest owing to the fact that the medium in which they are travelling is moving with a speed exactly equal and opposite to that of the velocity of propagation of the waves. Such waves can be observed where a rapidly flowing river passes over a large stone. A series of waves will be produced which, as far as the bed of the river is concerned, are at rest, but of course they are really travelling *up* stream with a speed equal to that of the water down stream.

PART II

SOUND

CHAPTER III

PRODUCTION AND PROPAGATION OF SOUND-WAVES

93. Nature of Sound-Waves.—That a body which is producing a sound is in a state of vibration can in general be at once perceived either by the sense of sight or touch. Further, as the amplitude of the vibrations of the body decreases so does the loudness of the sound decrease.

In order that we may perceive a *sound* as such, it is necessary that the body which is in a state of vibration shall be connected to our ear by an uninterrupted series of portions of matter, which portions of matter may be in the solid, liquid, or gaseous condition. In the absence of matter no sound can be communicated from the vibrating body to the ear, as can be proved by enclosing a bell, which can be struck by clock-work, within the receiver of an air-pump. As the air is exhausted the sound heard decreases, till when a fairly good vacuum is reached, although the hammer can still be seen striking the bell, yet no sound is heard. Hence we conclude that it is due to some disturbance in the air that the sound is able to travel out from the bell. Experiment shows that sound is in all cases communicated by *waves* produced in matter, and that these waves are longitudinal waves (§ 89). Let AB (Fig. 165) be the direction in which a sound-wave is travelling, say in air, and at a given instant *suppose* the air particles along AB are displaced from their mean positions to an extent represented by the ordinates of the curve AcDeB. It must be remembered that the displacements *actually* occur *along* the line AB, so that at the given instant the particle of which C is the *undisturbed* position is actually to the right of C by an amount equal to Cc. In the same way the particle at D is in its mean position, while the particle for which E is the mean position is displaced to the left by an amount Ee. Owing to the displacement of the particles a crowding together occurs which at the given instant is a maximum at D, and owing to this crowding together the pressure at D is increased. At

A and B in the same way the pressure is below the mean pressure, while at C and E the pressure has the mean value, that is the same value as exists before the wave occurred. The pressure variations at the different parts of the wave are shown by the ordinates of the curve $a'c'd'E'b'$. The curve $a''c''d''E''b''$ in the same way gives the *velocities* of the air particles at the different parts of the wave. Thus at the points A and B the displacement is zero, the pressure is a minimum, while the velocity is a maximum and towards the right. At C the displacement is a maximum to the right, the pressure has the mean value, and the velocity of the particle is zero. At D the displacement is zero, the pressure is a maximum, and the velocity is a maximum and towards the left, and so on for the other points.

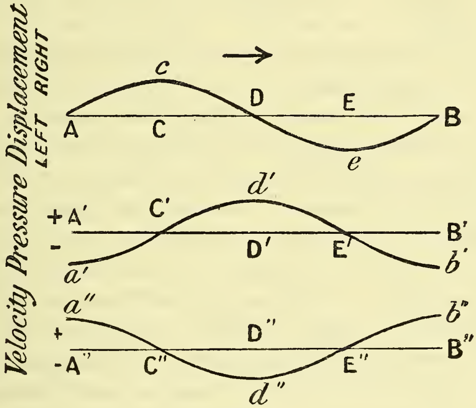


Fig. 165.

The distance from A to B is the *wave-length*, while Cc is the *amplitude* of the displacement of the air particles of the sound-wave, while D′d' is the amplitude of the pressure variations. The velocity with which any particular point of the wave, say the point D of maximum compression, travels forward is the velocity of the sound-wave. It is important to remember that the displacement of the particles are always in opposite directions on the two sides of a point of maximum or minimum pressure.

94. Velocity of Sound in Air and Water.—The fact that sound takes a very appreciable time to travel over a considerable distance is familiar to all. Thus, unless the storm is passing immediately overhead, there is a considerable interval between the instant when we see a lightning flash and that at which we hear the thunder, that is, the sound produced by the discharge.

Since the speed at which light travels is about 186,000 miles per second, the interval taken by a flash of light to travel over a few miles is

excessively small. Thus a direct determination of the velocity of sound in air can be made by observing the interval which elapses between seeing the flash of a distant cannon and hearing the report. Although the principle involved in this method of measuring the velocity of sound is very simple, the actual performance of the experiment involves considerable difficulty. Thus, as we shall see later, the velocity of the sound depends on the temperature of the air and on the humidity, and the determination of these quantities at *all* points between the gun and the observer is difficult. Further, what is required is the speed of the sound-waves in still air, while if the wind is in the same direction as the direction of propagation of the sound, the measured velocity will be too great, since the waves will be carried forward by the motion of the air. If the wind is against the direction of propagation of the sound, the velocity will in the same way appear to be reduced. Finally, there is a personal error produced by the observer, since a certain delay always occurs between a flash of light or a sound reaching us and the instant when we can, say, press a key to record the event. Taking this last cause of error first. If the delay in recording a signal received by the eye were the *same* as that when recording a signal received by the ear no error would be produced on the *interval* measured. However, this delay is not the same, and hence Regnault in his experiments replaced the human observer by a mechanical arrangement. He had two membranes, A and B, fitted with electrical contacts, so that when a sound-wave struck the membrane an electrical circuit was closed. One, A, was placed near the gun, and the other, B, at a measured distance away. Wires from A and B were taken to a recording drum, where by means of small electromagnets two pens were caused to record the instants at which the currents were produced. Now the time taken by the electric currents to flow through the connecting wires was negligible, and hence the time taken by the sound-waves to travel from A to B could be read off from the record on the drum. In order, as far as possible, to eliminate the effect of the wind, the signals were sent alternately from A to B and from B to A.

The numbers obtained by different observers for the velocity of sound in dry air at 0° C. vary between 330·6 and 331·9 metres per second, so that the value 331 metres per second, or 1086 feet per second, may be taken as fairly correct.

The velocity of sound in air is independent of the frequency of the waves, and as long as the amplitude is not too great the velocity appears to be independent of the amplitude. With very intense sound-waves, however, the velocity is a little larger than with waves of moderate amplitude.

The velocity of sound in water has been measured by a method similar to that described above, the values obtained being about 1730

metres per second at a temperature of 18° C. Here again the velocity increases slightly with increase in the amplitude of the waves.

The direct determination of the velocity of sound in solids presents such great difficulties that indirect methods, dependent on the measurement of wave-length and frequency, have to be employed. These, as well as similar methods which can be used in the case of gases, will be described later.

95. Method of Calculating the Velocity of Sound. Effect of Temperature on the Velocity of Sound in Gases.—As we have mentioned in § 89, the velocity, V, of a wave is connected with the elasticity E and the density D of the medium by the equation

$$V=\sqrt{\frac{E}{D}}$$

The value to give to the density in this expression is of course that corresponding to the temperature and pressure of the undisturbed medium, and presents no difficulty. The question as to the value to give to E requires further consideration, at any rate in the case of a gas. We have in § 40 seen that if the temperature of a gas is kept constant the elasticity is numerically equal to the pressure. Now since if we change the pressure, and hence the volume, of a gas heat will either be evolved or absorbed (§ 79), when a sound-wave passes through a gas heat will be developed at the points of the wave where a condensation exists, and absorbed at the points where a rarefaction exists. Thus the temperature of the gas will be raised at a condensation, and lowered at a rarefaction. Directly the temperature is raised, or lowered, in this way a flow of heat to the neighbouring portions of the gas will start, and if the condensations and rarefactions are sufficiently slow, very little actual difference of temperature will be produced owing to this flow of heat, and hence we may consider that the condensations and rarefactions occur isothermally, and hence E will be equal to P, the pressure. If, however, the condensations and rarefactions take place very quickly, the heat will not have time to flow to any appreciable extent from the hot portions, or to the cold portions, of the gas from the neighbouring particles, and hence the condensations and rarefactions will be almost adiabatic. Now in § 80 (equation 80) we saw that under adiabatic conditions the elasticity was γP, where γ is the ratio of the specific heat at constant pressure to that at constant volume. The question therefore arises whether in the case of sound-waves in a gas the changes of pressure are sufficiently slow to allow the temperature to remain constant, or are they sufficiently rapid to allow us to neglect the effect of transfers of heat. To answer this question we may calculate the velocity of sound in air under the two conditions. The density of air at 0° C. and under a pressure of 1013300 dynes per

square centimetre (one atmosphere) is 0·001293, and the value of γ is 1·41. Hence

$$\text{Isothermal velocity} = \sqrt{\frac{1013300}{0{\cdot}001293}} = 27995 \text{ cm. per second.}$$

$$\text{Adiabatic velocity} = \sqrt{\frac{1013300 \times 1{\cdot}41}{0{\cdot}001293}} = 33319 \text{ cm. per second.}$$

As the actual velocity is 33100 centimetres per second, we infer that the adiabatic hypothesis is more nearly correct. The isothermal hypothesis was made by Newton in his original calculation of the velocity of sound, and Laplace was the first to point out that the adiabatic elasticity was the one which ought to be employed.

In the case of liquids and solids the changes in volume which accompany the passage of sound-waves are so small that no appreciable variations in temperature occur, and hence the isothermal elasticity gives a correct result for the velocity. Hence for a liquid the bulk modulus (§ 48) can be used in calculating the velocity of sound. When dealing with the velocity of sound along rods we may use Young's modulus.

Effect of temperature on the velocity of sound in a gas.

If p_0, v_0 are the standard pressure and the volume of unit mass of a gas at this pressure and at 0°, while p and v are the corresponding quantities at a temperature t, we have (§ 63)

$$\frac{pv}{1+at} = p_0 v_0$$

where a is the coefficient of expansion of the gas and has the value 0·00366. If V_t is the velocity of sound at the pressure p and temperature t we have, since the density is equal to $1/v$,

$$\begin{aligned} V_t &= \sqrt{\gamma p v} \\ &= \sqrt{\gamma p_0 v_0 (1+at)} \end{aligned}$$

But the velocity at a pressure p_0 and 0° C. is given by $V_0 = \sqrt{\gamma p_0 v_0}$ Hence

$$V_t = V_0 \sqrt{1+at} \quad . \quad . \quad . \quad . \quad (110)$$

But since a is small, $\sqrt{1+at} = 1 + \frac{at}{2} = 0{\cdot}00183t$.

Thus
$$V_t = V_0(1 + {\cdot}00183t) \quad . \quad . \quad . \quad (111)$$

It will thus be seen that changes in pressure alone do not affect the velocity of sound in a gas, the reason being that a change in pressure is accompanied by a proportional change in density (Boyle's law), so that the expression $\sqrt{\frac{P\gamma}{D}}$ is unaltered. Increase of temperature, however, causes the velocity to increase.

96. Pitch. Musical Scale.—Sounds can be roughly divided into two classes—noises and musical notes. These are both due to waves, but in the case of a noise the character of the waves, *i.e.* the frequency, amplitude, &c., is very complex, while in the case of a musical note the character of the waves is in general much more simple.

The most noticeable distinction between musical notes is that of *pitch*, the pitch depending on the period of the waves that constitute the sound. A high pitch corresponds to a small period that is a great frequency, while a low pitch corresponds to a small frequency. Hence the pitch of a note can be defined by the frequency, that is by the number of vibrations per second. Thus the pitch of the note produced by a vibrating body is equal to the number of vibrations the body makes per second. The pitch of the sound-waves sent out by the vibrating body is the number of waves which pass a given spot in a second, and, as has been shown in § 89, this number of waves is equal to the number of times a particle of the medium vibrates, or since the particles of the medium in the immediate vicinity of the vibrating body are moved by this body, the number of vibrations the particles make in a given time is equal to the number of vibrations made by the body. Thus we obtain the same value for the pitch of the sound whether we measure the frequency of the vibrating body which produces the note or that of the waves produced in the medium surrounding the vibrating body.

The interval between two notes.

If we have two notes of which the frequencies are n_1 and n_2, then if the ratio of n_1 to n_2 is expressible by two *small* integers such as 1 to 2 or 2 to 3, then even an untrained ear is able to appreciate the fact that as long as the *ratio* of the frequencies remains the same the two notes are related to one another in some way, which remains unaltered although their pitch may be altered. Thus notes of which the frequencies are 256 and 384 are, as far as the ear is concerned, related to one another in the same manner as two notes of which the frequencies are 512 and 768. In each case the ratio of the frequencies is 2/3, and this ratio is called the *interval* between the two notes.

The musical scale.

Starting with a note of any arbitrary frequency, we can select a series of other notes of which the intervals from the original note are quite definite, and these notes are recognised by all civilised nations as constituting a series, and form what is called a *musical scale*. The question as to why one particular set of intervals should be recognised in this way would lead beyond the scope of this book. It may, however, be mentioned that probably in the case of every individual training plays an important part, in that from infancy all music which is ever heard employs the ordinary scale. Other more truly physical reasons will be referred to later.

The note from which the scale starts, called the *tonic*, is generally indicated by the letter *C* or the name *do*, while the remaining notes are indicated by the letters and names shown in the following table, in which are also given the intervals between each note and the tonic, and the intervals between consecutive notes :—

Number of note	1	2	3	4	5	6	7	8
Letter notation	C	D	E	F	G	A	B	c
Name	do	re	mi	fa	so	la	si	do′
Interval between note and tonic	1	$\frac{9}{8}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{15}{8}$	2
Interval between consecutive notes		$\frac{9}{8}$	$\frac{10}{9}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{9}{8}$	$\frac{16}{15}$
Series of lowest integers which form scale	24	27	30	32	36	40	45	48

Special names are given to certain intervals. Thus two notes which have the same pitch, so that the interval is unity, are said to be in *unison.* If the interval is 2, so that the frequency of one note is double that of the other, the notes are said to differ by an *octave,* the name being chosen because the first and eighth notes of the scale have an interval of 2. The interval of 3/2, *i.e.* that existing between the first and fifth notes, is called a *fifth.* That of 4/3, between the first and fourth notes, is called a *fourth,* and the interval of 5/4, between the first and third, is called a *third.* In the same way the interval of 5/3 is called a *sixth,* while the intervals of 9/8 and 10/9 are called a *major tone* and a *minor tone* respectively. The interval 16/15 is called a *limma.*

In addition to the notes of the scale given above, use is made in music of additional notes, which are derived from the above by either raising or lowering the pitch of each note by 25/24 or 24/25, this interval being called a *diesis* or *semitone.* When the pitch is raised by a semitone the note is said to be *sharp,* while if the pitch is lowered the note is said to be *flat.*

Nothing has been said above about the frequency of the tonic, for this is quite arbitrary, and has differed at different times, and does differ in different countries.

The musical scale is not confined to a single octave, but, starting with the octave of the original tonic as a new tonic, a new series of notes is obtained, and so on, the frequency of each note in any octave being twice that of the corresponding note of the octave below and half of that of the corresponding note in the octave above.

97. Measurement of the Pitch of a Tuning-Fork. Stroboscopic Method.—The standards of pitch usually employed are U-shaped metal rods called tuning-forks. The fork A (Fig. 166) is generally

screwed by the tang to a wooden box B, which is of such a size as to resound (§ 103) to the note given by the fork.

One method by which the frequency of a fork can be measured is to tune a string, which is stretched by a known force, to unison with the fork, and to calculate the pitch of the note given by the string by the formula given in § 101. This method does not, however, admit of any great accuracy. A better method consists in attaching a fine bristle to one prong of the fork, and causing this bristle to trace a wavy line on a smoked drum, which is rotated at a fairly rapid rate. A small electromagnet works another style, which traces a second line on the drum alongside that due to the fork. The current of the electromagnet is made at equal intervals by the pendulum of a standard clock, so that the line traced on the drum is broken, and the time interval between these breaks is known. Hence, by counting the number of

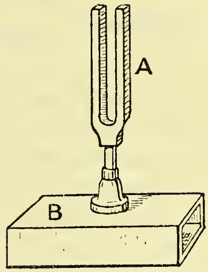


FIG. 166.

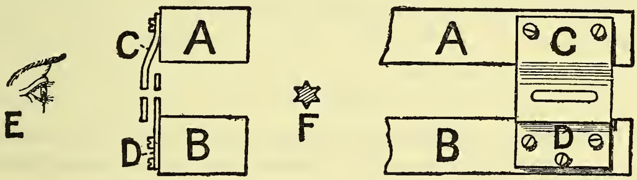


FIG. 167.

vibrations of the fork which occur between two given time-marks, the frequency of the fork can be obtained. Another method of considerable importance is that known as the stroboscopic disc method. Two thin and light pieces of card or metal foil C, D (Fig. 167) are attached to the prongs A, B of the fork. Each of these cards is perforated by a slit, and these slits are so placed that, when the prongs of the fork are at rest, the two slits are opposite each other, so that an eye placed at E can see through. When the fork is sounding, it will only be possible to see through when the prongs are passing through their position of rest, and hence an object placed at F will be seen intermittently, the interval between two views being equal to the interval between two

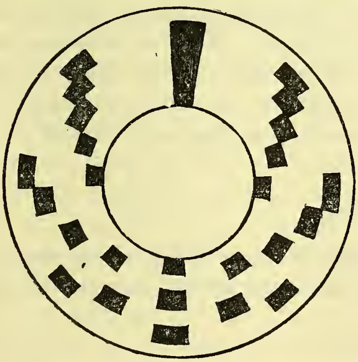


FIG. 168.

consecutive passages of the prongs through their positions of rest, that is, at intervals equal to half the period of the fork. If the object at F is a disc on the face of which are painted a number of rings of equidistant dots, as shown in Fig. 168, then, if this disc is in rotation, and, during the time which elapses between two views through the slits, a dot in any one of the rings has just had time to take the position occupied by the preceding dot when the disc was seen before, this ring of dots will appear at rest, for the eye cannot distinguish between the different dots, and whenever it sees the disc, the dots on the ring considered are as a whole in the same position.

Stroboscopic disc.

When a ring of dots appears at rest it means that each time the disc is seen each dot has moved on to the position occupied by the preceding dot at the last glimpse. Thus the number of dots which pass any given point in a given time is equal to the number of glimpses of the disc which are obtained in this time. If there are m dots on the ring and the disc makes n revolutions in a second, the number of dots which pass is nm. Since a glimpse is obtained every half-vibration of the fork, if N is the frequency of the fork, there will be $2N$ glimpses per second, and hence

$$2N = mn \quad . \quad . \quad . \quad . \quad (112)$$

Thus if the number of revolutions of the disc in a given time is measured by means of a revolution-counter, n can at once be calculated, and hence the frequency of the fork obtained by the above equation.

If the speed of rotation of the disc is slightly greater than that for which a given row of dots appears stationary, then each time a glimpse of the disc is obtained a given dot will have slightly passed the position occupied by the preceding dot at the preceding glimpse. Hence the ring of dots will appear to be rotating slowly in the same direction as that in which the disc is rotating. During the time the ring of dots appears to advance through the distance between two dots, one more dot will have passed any point than the number of glimpses. Hence if x dots appear to pass in a second, we have

$$2N = mn - x.$$

Similarly, if the disc is rotating too slowly the ring of dots will appear to turn slowly in an opposite direction to that in which the disc rotates, and in this case we have

$$2N = mn + x.$$

The above principle may often be used to study the motion of a vibrating body, for if we have a rotating disc furnished with one or more radial slots, and so adjust the speed of the disc that the interval which

elapses between successive glimpses, when we look at the vibrating body through the slots, is slightly greater than the period, the body will appear to go through its vibratory motion quite slowly. Another application of the same method is the following. Suppose a speed-indicator or revolution-counter is connected to a rotating shaft by a belt. If the diameters of the pulleys on the shaft and on the speed-indicator are exactly the same and there is no slip the reading given will correspond to the speed of the shaft. If, however, the above conditions are not exactly satisfied a correction will have to be applied to the readings. By mounting a disc in which is cut a radial slot on the axis of the speed-indicator, and painting a dot on the pulley or some other part carried by the shaft, and looking through the slot at the dot we can at once see whether a correction is necessary. If the disc and shaft revolve at exactly the same speed the dot will always be seen in exactly the same place. If, however, the speeds are not exactly the same the dot will appear to rotate either in the same or the opposite direction to that of the shaft. The time the dot takes to complete a revolution gives the time the shaft gains or loses, according to the direction of the dot's apparent motion, one whole revolution on the speed-indicator.

98. Reflection and Refraction of Sound.—Sound being conveyed through the air by waves, when these waves meet a solid obstacle they are reflected in the manner described in § 92. A familiar example of the reflection of sound is afforded by an echo, which is due to the reflection of the sound-waves from some large surface, such as a cliff or the side of a house.

The sounding-board placed over the head of the speaker in large halls is another application of the reflection of sound. It consists of a reflecting surface placed so as to reflect those sound-waves that strike it down towards the audience. Hence the waves, that would otherwise spread up to the roof and be irregularly reflected there, are directed downwards, and assist in making the speaker audible.

In the case of speaking-tubes, the sound-waves, instead of spreading out in spheres, as they would do in the open air, are, by reflection at the sides of the tube, confined within the tube, so that they travel forward with comparatively small decrease in amplitude, the wave-front remaining of the same cross-section throughout. A similar effect is produced when a watch is held against one end of a long wooden rod, and the ear is held against the other end. The ticking of the watch can be heard almost as clearly as if it were held close to the ear. The reason is that the sound-waves in the wood, when they reach the bounding surface between wood and air, are almost entirely reflected, and thus the wave proceeds down the rod without much of the energy escaping into the surrounding air.

The difficulty of hearing sounds at a distance on certain days is supposed to be due to the fact that on such acoustically opaque days there exist columns and layers of air at different temperatures, and that the sound-waves get partly reflected at each passage from air at one temperature to air at another. Such reflection must occur, for the velocity with which the sound-waves travel will be different if the temperature of the air is different, and whenever a wave passes from one medium to another, in which it moves with a different velocity, it is partly reflected at the boundary between the two media.

Musical echo.

Another example of reflection is the musical note occasionally heard when a sharp sound, such as a clap, is produced near a wooden paling or a flight of steps. The clap produces a single wave, and a portion of this wave is reflected when it strikes each of the palings, so that if the individual palings are at different distances a *succession* of reflected waves will be produced which, reaching the observer at equal intervals, will produce the sensation of a train of waves, *i.e.* a musical note will be heard, the pitch of which depends on the interval between the individual reflecting surfaces.

When sound-waves which are travelling in one medium reach a surface separating this medium from another they are partly reflected and partly transmitted. In general the direction of propagation in the second medium is different from that in the first, that is, the waves are refracted. The subject of refraction of sound-waves is of very little interest, and hence the study of refraction is postponed till we deal with light-waves where the phenomena are of great importance.

Refraction caused by wind.

There is a change of direction due to wind which has some importance in sound. This wind refraction is due to the fact that in general the velocity of the wind is less near the surface of the earth than at some distance up. Thus suppose we have a plane sound-wave, in which the wave-front is vertical, moving against the wind. Now the sound-wave will travel at the same speed through the air, but, owing to the contrary motion of the air, the distance moved through by the wave-front *relative to the earth* will be greater near the surface of the earth than higher up. The wave-front will therefore become inclined, the top lagging behind the bottom, and, since the motion of the wave is at right angles to the wave-front, the direction of motion of the wave, instead of being parallel to the surface, will be inclined upwards. Thus the sound-waves may pass over the head of an observer who is on the windward side of the place where the sounds originate. When the sound is travelling with the wind the opposite effect is produced, the waves being refracted downwards. This effect accounts for the greater distance sounds can be heard when the sound is moving with the wind, than when the sound is moving against the wind.

Refraction of sound-waves is also occasionally produced by the varying temperature of the air. During the day the air near the ground is generally at a higher temperature than that above, and since the velocity of sound increases with rise of temperature an effect is produced similar to that caused by an opposing wind. At night, when the temperature often increases as we go upwards, an opposite effect is produced, and hence sounds can be heard at greater distances than during the day.

99. Doppler's Principle.—Suppose that at a point A there is a body which is emitting a note, of which the frequency is n. Owing to the action of the sounding body there will be a succession of waves produced in the surrounding air, and the frequency of these waves will also be n. Hence, if an observer is at a point B, n waves will reach his ear in each second, and he will hear a note of pitch n. Now suppose the observer approaches the sounding body, then in each second he will now receive more than n waves, for in addition to the n waves which would reach his ear, suppose he had been stationary, he will have met a certain number of extra waves in each second, for at the end of the second he is nearer the sounding body than he was at the commencement, and his ear will have received the waves which, at the commencement of the second, occupied this space. The result is that as he now receives more than n waves per second, the pitch of the note he hears will be higher than n. If, instead of approaching the sounding body, he moves away from it, then in the same way the pitch of the note heard will appear lower than n. If the observer remains at rest, but the sounding body approaches, then the effect will be the same as if the observer approached a stationary sounding body, and the note heard will be of a higher pitch than that produced by the body; while if the sounding body is receding, the note heard will be lower.

This apparent change of pitch, owing to the relative motion of the sounding body and the observer, is called the Doppler effect, and the explanation which we have given is called the Doppler principle.

The Doppler effect can often be observed in the case of the note given by a steam-whistle sounding on an engine which is passing through a station at a rapid rate. The note heard when the engine is approaching the observer is very markedly of a higher pitch than that heard when the engine has passed and is travelling away from the observer.

The change in pitch produced by the relative motion of the observer and sounding body can readily be calculated. Suppose that the sounding body has a frequency of n vibrations per second, and that the observer is approaching it with a velocity v, the velocity of the sound being V. If the observer were at rest, he would receive n waves per second. If l is the wave-length of the sound, then in a space v there will be v/l waves. Hence, since the observer traverses a space v in one second, his ear will

pick up, owing to his motion alone, v/l waves, so that the total number of waves received in a second will be $n+v/l$. But $V=nl$ (§ 89), so that the pitch of the note heard is $n+nv/V$, or $n(1+v/V)$. If the observer had been travelling away from the sounding body, then in each second his ear would have failed to pick up v/l waves, and the pitch of the note heard would be $n(1-v/V)$.

While Doppler's principle is of no practical importance in sound, in light, however, it has enabled estimates to be made of the speed at which the solar system is moving towards or away from certain of the stars.

100. Interference of Sound-Waves. Beats.—The fact that two series of waves of the same period can at some points destroy each other's effect, that is, produce interference (§ 91), can be shown by holding a vibrating tuning-fork near the ear, or over the mouth of its resonator, and slowly turning the fork about the stem as an axis. Four times in each revolution no sound will be heard. When the fork vibrates, the prongs move alternately towards and away from one another. When the prongs are approaching one another a condensation is started in the space between the prongs. At the same time, however, each prong produces at E and F (Fig. 169) a rarefaction. When the prongs separate, a rarefaction is produced between the prongs, and condensations are produced at E and F. Hence each prong starts two sets of waves, and these trains of waves start in opposite phases, so that at all points along the lines NAM and N′BM′ a condensation and a rarefaction always arrive simultaneously, and these, by interfering, produce rest, *i.e.* silence.

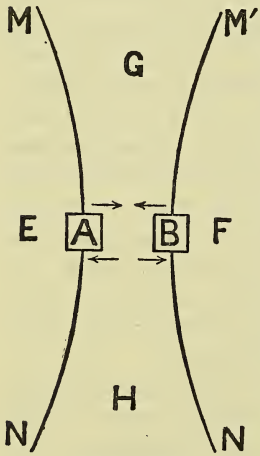


Fig. 169.

When a high-pitched whistle is sounded before a plane reflecting surface, the direct and reflected waves interfere and produce stationary waves as described in § 92. The nodes and loops of these stationary waves can be determined by moving a glass tube, connected to the ear by rubber tubing, along the normal to the surface, when maxima (nodes) and minima (loops) of sound will be observed. The nodes are positions of minimum displacement of the air particles, but for the reasons given in § 93 these are the positions of *maximum change in pressure*, and it is when the change in pressure at the end of the exploring tube is a maximum that the strongest waves are produced in this tube. At a loop the air particles sweep backwards

Stationary sound-waves in air.

and forwards across the end of the tube, and produce very little disturbance in the tube itself. The position of the nodes can also be determined by means of a sensitive flame. This consists of a pin-hole burner supplied with illuminating gas under pressure. When no sound-waves strike the burner the gas burns with a long steady flame. The incidence of sound-waves, particularly if they are of high pitch, cause the flame to shorten and roar.

The above experiment enables us to measure the wave-length of a sound-wave, for the distance between consecutive nodes is equal to half a wave-length.

Beats.

When two forks of nearly the same pitch are sounded together the resultant sound waxes and wanes in a very characteristic manner, and *beats* are said to be produced. The production of beats is due to the phases of the two sets of waves sent out by the forks being alternately the same and opposite. This effect has already been considered in § 85, where we were dealing with the resultant of two S.H.Ms. of nearly equal period.

Let one fork (A) make x vibrations, while the other (B) makes $x+1$. If then we start when the two are in the same phase, the phase of the fork B will gain on that of A, till at the end of x vibrations of A, B will have made $x+1$ vibrations, and so they will again be in the same phase, and the sound will be a maximum. Let the frequency of the fork A be nx, and hence that of B $n(x+1)$. Now from one maximum of sound to the next A makes x vibrations, so that the number of maxima in a second will be n, or there will be n beats per second. But the difference between the frequencies of A and B is $n(x+1)-nx$ or n, so that the number of beats per second is equal to the difference in the frequencies of the two forks.

Thus by counting the number of beats made by two notes in a given time we can deduce the number of beats per second, and this gives the difference in the frequencies of the two notes. Since, however, we do not know which is of the higher pitch, an auxiliary experiment is necessary. This consists of weighting the prongs of *one* of the forks with very small pieces of wax and again determining the number of beats produced per second. The addition of the wax reduces the frequency of the fork, and hence if the frequency of the beats increases, so that the difference in pitch has been increased, we infer that the weighted fork had the lower pitch. If, on the other hand, the frequency of the beats is decreased by the loading, the loaded fork must have possessed the higher pitch.

CHAPTER IV

VIBRATIONS OF STRINGS, RODS, AND COLUMNS OF GAS

101. Vibration of Strings.—If a thin string of which the mass of unit length is m is stretched by a tension T, and is struck or plucked, a wave will travel along the string. The velocity, v, with which this wave travels will be independent of the form of the wave, and is given by

$$v=\sqrt{\frac{T}{m}}$$

When the wave reaches the fixed end of the string it will be reflected in the same way that a water-wave is reflected when it strikes a wall. There is, however, this important difference in the two cases. As we have seen in § 92, when a crest of a water-wave strikes a wall, a crest starts back in the reflected wave, so that at the reflecting surface the direct and reflected waves are always in the same phase. In the case of the string, on the contrary, the direct and reflected waves are in opposite phases at the point of reflection, so that if a crest travels up to the fixed end a trough travels back. The reason for the difference lies in the different conditions which hold at the point of reflection in the two cases. In the water-wave the up and down displacement of the water particles is unconstrained at the surface of the wall, but in the case of the string, at the fixed end the movement must of necessity be zero. As a result, while in the case of the stationary vibrations set up when a series of water-waves are reflected at a wall a loop exists at the surface of the wall, when stationary vibrations are set up by sending a train of waves along a string so that they are reflected at a fixed end, the end is a node.

If we have a string AB fixed at both ends, and pluck it at the middle point C, that is draw it upwards and release it, a wave will travel away from C in both directions, and if v is the speed of propagation and l the length of the string, the two waves will reach the ends A and B in a time $l/2v$. They will be reflected at A and B respectively with change of phase, so that downward displacements will be reflected back, and these will reach the middle point at a time $l/2v$ after reflection, *i.e.* at a time l/v after the start. The reflected waves will strengthen each other and combine to produce a downward displacement, and after each has traversed the whole string they will again be reflected at the ends, and

will finally come together at the mid point at a time $2l/v$ from the start, and will combine to give an upward displacement. The whole cycle of operations is then repeated, and as a result the string will swing backwards and forwards transversely about its position of rest, the period of a whole vibration being $2l/v$. But $v=\sqrt{\frac{T}{m}}$, so that the period t is given by

$$t=\frac{2l}{v}=2l\sqrt{\frac{m}{T}}$$

or if the frequency is n,

$$n=\frac{1}{t}=\frac{1}{2l}\sqrt{\frac{T}{m}} \quad . \quad . \quad . \quad . \quad (113)$$

We might have derived this formula by considering a string fixed at both ends and vibrating transversely as a case of stationary vibrations, with the two ends as nodes, when we should have that the length of the string could be half the wave-length. But (§ 89) the velocity is equal to the wave-length l' multiplied by the frequency. Hence

$$v=nl'=2nl=\sqrt{\frac{T}{m}}$$

Equation (113) above may be written in a somewhat different form. If r is the radius of the string, and D is the density of the material of which it is composed, $m=\pi r^2 D$. Hence

$$n=\frac{1}{2l}\sqrt{\frac{T}{\pi r^2 D}} \quad . \quad . \quad . \quad . \quad (114)$$

Overtones of string.

In addition to the mode of vibration considered above, which gives the lowest note which a string can produce and which is called the *fundamental*, it is possible to cause a string to vibrate in such a way that 1, 2, 3, &c., nodes, or points of permanent rest, exist between the fixed ends. These additional modes of vibration are called overtones, and the fundamental mode and the first two overtones are shown in Fig. 170. Since for the first overtone the distance between consecutive nodes is half the length of the string (the fixed ends are nodes) it follows at once from equation (113) that the frequency is twice that of the fundamental. In the same way the frequency for the next overtone is three times that of the fundamental, and so on. When the frequencies of the fundamental and the overtones are as the numbers 1, 2, 3, 4, &c., the overtones are called the *harmonics* of the fundamental.

We have above supposed that the string vibrates in only one mode at a time; we may, however, have a number of different modes coexisting,

the resultant motion being obtained by compounding the individual modes. Thus the sound produced by a string may be a compound note which consists of the fundamental and the overtones. If we strike or pluck a string at any given point, this point cannot possibly be a node, and hence any of the overtones which require this point as a node will not be produced. Thus when a string is plucked at the mid point the first overtone is never produced.

Transverse vibrations of rods.

In the case of a string the force of restitution when the string is displaced is entirely due to the tension, at any rate so long as the string is not very stiff and short. We have now to consider the transverse vibrations of a rod, where no tension exists but the force of restitution is entirely due to the stiffness.

If the rod is held in a clamp at one end, the fundamental form in

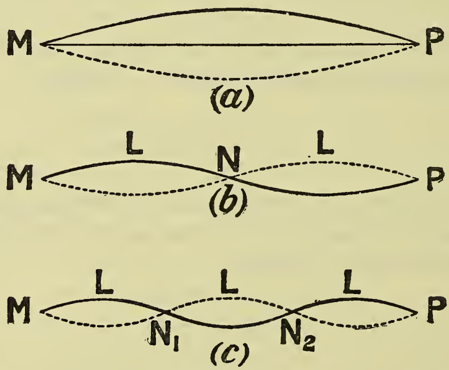


FIG. 170.

which it can vibrate is shown at (*a*), Fig. 171, where there is a single node, and that at the fixed end.

The manner in which a rod clamped at one end vibrates, when sounding its first and second overtones, is shown at (*b*) and (*c*). If the frequency of the rod when sounding its fundamental is taken as unity, then the frequencies of the overtones are 6·27, 17·55, 34·39, 56·84, &c. In this case it will be observed that the overtones are not the harmonics of the fundamental tone.

If the rod, when vibrating as in (*b*), Fig. 171, instead of being clamped at B, were prolonged, and were simply supported at N and B, we should have the case of a rod free at both ends, and vibrating in its fundamental form as shown at (*a*), Fig. 172. The mode of vibration for the first overtone is shown at (*b*), and is such that there are three nodes.

The relative frequencies of the fundamental and of the overtones for a rod free at both ends are 1, 2·76, 5·40, 8·93, 13·35, &c.

The consideration of the connection between the dimensions of a rod and the frequency of its fundamental tone is beyond the scope of this work. It is, however, interesting to note that, other things being the same, the frequency varies inversely as the square of the length, and in the case of rectangular rods, directly as the thickness, or as the radius in cylindrical rods.

In a case of a string or rod, in addition to the transverse vibrations considered above we may have longitudinal vibrations in which the particles move backwards and forwards parallel to the length of the string or rod.

Longitudinal vibrations of strings and rods.

The frequency of the longitudinal vibrations of a string is independent of the tension with which the string is stretched.

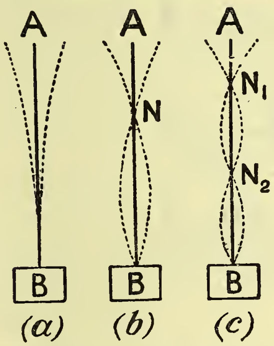


FIG. 171.

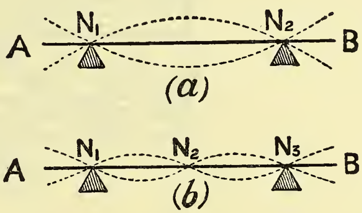


FIG. 172.

For when a particle of the string is displaced from its mean position, the force with which it tends to return to its undisplaced position depends on the stress caused by the *displacement* from this position, and this stress is by Hooke's law (§ 73) independent of any previously existent stress which affects the particle under consideration and the other particles equally. Thus the velocity with which a longitudinal disturbance travels in a string is independent of the tension, and depends only on the elasticity and density of the material of the string. When the string is giving its fundamental, there will be a node at each end and a single loop in between, so that the wave-length will be equal to twice the length of the string. Hence, since $v=nl$, where v is the velocity of sound in the material of the string, n is the frequency of the note produced, and l is the wave-length, we get $v=2nL$, where L is the length of the string.

In the case of a rod clamped at the middle and vibrating longitudi-

nally, there must be a node at the centre where the rod is held, and the ends of the rod must always be loops. Hence when the rod is sounding its fundamental there will be a loop at either end, and a single node, namely that at the centre. The wave-length of the sound in the rod will therefore be equal to twice the length of the rod.

Since, as in the case of the longitudinal vibrations of a string, $v=2nL$, if we measure the frequency n of the note given by a rod or string of length L when vibrating longitudinally, we can immediately calculate the velocity of sound in the material of which the rod or string is composed, and it is in this way that the velocity of sound in solids is usually obtained.

102. Vibrating Columns of Gas.—The column of gas, say air, enclosed in a tube can be caused to vibrate longitudinally in a manner strictly analogous to that of the longitudinal vibrations of rods. Two

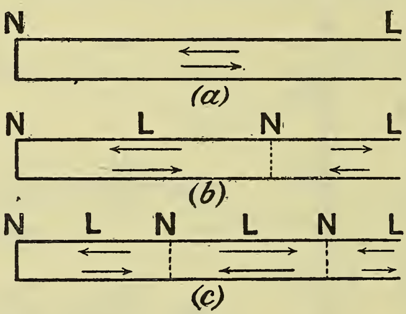


Fig. 173.

cases have to be considered, namely, that in which the tube is open at both ends, and that in which the tube is closed at one end.

In the case of vibrating columns of air, at the nodes, which are points where the air particles are at rest, there will be maximum change of pressure, for the particles will alternately be crowded together and separated at these points. The loops, on the other hand, will be places of maximum motion, but of minimum change of density and pressure.

In the case of a closed pipe, there can be no motion of the air particles which are in immediate contact with the closed end, so that the closed end must always be a node. At the open end, where the air column communicates with the external air, the changes of density can only be very small, so that the open end may for the present, at any rate, be regarded as a loop. Hence the fundamental is produced when the air column vibrates, as at (*a*), Fig. 173. The wave-length will be equal to four times the length of the pipe, for it is always equal to four times the distance between a node and the adjacent loop. The first

overtone is produced when there is one node besides that at the closed end, as shown at (*b*), while the second overtone is produced when there are two additional nodes, as at (*c*). The air particles on the two sides of a node are always moving in opposite directions, and when a condensation is taking place at one node, a rarefaction is taking place at the adjacent nodes. The wave-length at (*b*) is equal to twice the distance between consecutive nodes, that is, is equal to $\frac{4}{3}L$, where L is the length of the pipe. Thus the wave-lengths of the fundamental and of the overtones of a closed pipe are

$$4L, \frac{4L}{3}, \frac{4L}{5}, \frac{4L}{7}, \text{ \&c.}$$

Since the velocity of sound in the air is the same in all cases, and $v = nl$, the frequencies of the fundamental and of the overtones are inversely

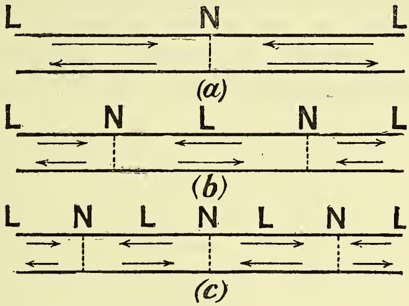


FIG. 174.

proportional to the wave-lengths, so that, if the frequency of the fundamental is taken as unity, the frequencies of the fundamental and of the overtones are 1, 3, 5, 7, &c. In this case, therefore, only the odd harmonics of the fundamental are present in the overtones.

In the case of a pipe open at both ends, there must be a loop at each end, and the fundamental is given when there is a single node produced at the middle, as shown at (*a*), Fig. 174. In this case the wave-length is equal to twice the length of the pipe. The modes of vibration corresponding to the first two overtones are shown at (*b*) and (*c*). It will be seen that the wave-lengths of the fundamental and overtones are equal to $2L, \frac{2L}{2}, \frac{2L}{3}, \frac{2L}{4}$, &c., or in the ratio of $1 : \frac{1}{2} : \frac{1}{3} : \frac{1}{4} : \ldots$ Hence the frequencies are in the ratio of

$$1 : 2 : 3 : 4 : \ldots$$

So that in the case of an open pipe all the harmonics of the fundamental are produced by the overtones.

If a hole is made in the side of the pipe no vibration can exist which requires a node at this point, for a node corresponds to a point of maximum change in pressure, and the hole communicating with the external air prevents any appreciable change of pressure taking place at that point. On the other hand, the presence of the hole is favourable to the formation of a loop at the point. Thus if a hole is made at the point L, Fig. 173 (*b*), the pipe will naturally give its first overtone. If, however, the hole is made at N in (*b*) the pipe will give the second overtone, that is, will vibrate in the mode shown at (*c*). This is the principle on which the flute works, the holes in the side being adjusted so as to give the overtones required to produce a musical scale.

Since the frequency of the fundamental of a closed pipe is $v/4L$ and that of an open pipe is $v/2L$, where v is the velocity of sound in the gas in the pipe, it is evident that the pitch in either case depends not only on the nature of the gas but also on the temperature. By means of the expression given on page 256, the increase in pitch produced by a rise in temperature for any given pipe can at once be calculated.

CHAPTER V

RESONANCE AND MAINTENANCE OF VIBRATIONS

103. Resonance.—When a vibrating body, such as a tuning-fork, is held near the end of a pipe, waves are produced in the gas in the pipe which are forced vibrations (§ 88), having the same frequency as that of the fork. If the natural frequency of one of the modes of vibration of the gas is the same as that of the fork, the amplitude of the forced vibrations set up is large, that is, we get resonance, and the column of gas is said to act as a resonator.

The chief function of resonators in acoustics is to strengthen the amount of the sound, of the particular pitch to which they respond, which is radiated through the surrounding air, that is, to increase the amplitude of the waves produced in the external air by means of which the sound is heard. As an example of this action of a resonator, we may take the case of the increase in the loudness of the note given by a tuning-fork by means of a resonator. From the fact that the prongs of a fork have not a very great area, they are not capable of setting any great quantity of the surrounding air in violent vibration, for the air on the side towards which the prong is moving can slip round the edge of the prong, and thus partly fill up the rarefaction which is being produced on the other side of the prong. In addition, the interference which takes place between the waves emitted from the two prongs (§ 100) reduces the intensity of the motion produced in the surrounding air. If, however, the sounding-fork is held near the open end of a closed pipe, of which the natural period is equal to that of the fork, that is, its length is equal to $\frac{1}{4}$ of the wave-length in air of the note given by the fork, then this pipe will be in resonance with the fork, and the column of air within the pipe will be set into vibration. Now if the open end of the pipe is not very small, the vibrations of the air inside will, at the open end, set the external air into vibration much more powerfully than the fork alone did. The result is that on bringing such a resonator near a sounding-fork, the intensity of the sound heard is very much increased. The same kind of effect can be easily noticed in the case of the vibrations of a string. Here, again, the surface of the vibrating body is small, so that it is incapable of setting any great mass of air into vibration, and in addition the waves produced on the two sides of the string are in opposite phase, for when a condensation is being produced on one side

S

of the string a rarefaction is being produced on the other side, and these two sets of waves interfere. By holding a pipe near, of which the natural period agrees with that of the string, or, what is better, connecting one end of the string to the walls of the pipe, so that the vibrations of the string are communicated to the walls, and by them to the air contained within the pipe, the column of air is set into vibration, and communicates its vibrations to the external air much more powerfully than the string alone is capable of doing.

The above examples of the strengthening of the sound produced by a vibrating body, which are cases of true resonance, must not be confounded with others where the vibrating body is able to set up *forced* vibrations in a body, and the increase in the loudness is due to these forced vibrations. Thus when the stem of a vibrating tuning-fork is pressed against the wooden top of a table, the loudness of the sound produced is greatly increased. In this case the fork sets the table into forced vibrations which have the same frequency as the fork, and the

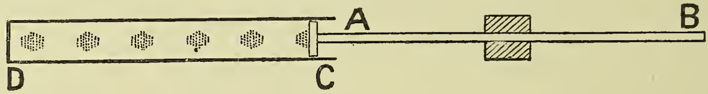


Fig. 175.

table, on account of its large surface, is able to produce violent sound-waves in the air. That this is not a case of resonance is shown by the fact that the table acts equally well in the case of forks of all frequencies.

Since the resonator obtains the energy necessary to set it into vibration from the sounding body, and since the increased loudness of the sound emitted when the resonator is present means that more energy is being given out to the external air, it follows that the sounding body must lose its energy of motion more rapidly when a resonator is present than it does when no resonator is present. That this is so can be easily shown by sounding a tuning-fork first without the resonance-box belonging to it, and then with the box, when it will be found that the vibrations of the fork continue for a much longer time without the resonator than they do with it.

Kundt's method of measuring the velocity of sound in a gas.

Kundt has applied the method of forced vibrations set up in a column of gas to measure the velocity of sound in a gas. His arrangement consists of a rod AB (Fig. 175) clamped at its mid point so that one end projects into a tube CD, the end of the rod being provided with a light piston which fits loosely into the tube. On causing the rod to vibrate longitudinally, the piston will vibrate backwards and forwards and will set up vibrations in the air contained in the tube CD. The waves in the air in the tube will be reflected from

the end D of the tube, and the direct and reflected waves will set up stationary vibrations in the air. If we suppose that the tube is closed at D, this point will be a node, and there will be a series of nodes along the tube at distances equal to $l/2$ from one another, where l is the wave-length in the gas which fills the tube of the tone having a frequency equal to that of the rod. If the position of the end of the rod is adjusted so that the piston B is at a loop of these stationary vibrations, resonance will occur and the motion of the piston will have its maximum effect in increasing their amplitude, and they will be so intense that if a light powder, such as lycopodium or cork filings, be strewn inside the tube, it will, by the vibration of the air or other gas, be collected in very characteristic transverse ridges at the loops. The explanation of the formation of these ridges is beyond the scope of this work, so we must content ourselves with referring the reader who wishes to pursue the subject to Rayleigh's *Sound*, vol. ii. p. 47. By measuring the distance between consecutive loops, we obtain the value of $l/2$ for the note produced by the rod in the gas, and this represents the space traversed by a sound-wave in the gas during the time the rod makes half a vibration. If n is the frequency of the rod, this will also be the frequency of the vibrations in the gas, so that if v is the velocity of sound in the gas, we have $v = nl$, or $v = 2nx$, where x is the distance between two of the loops in the tube. If the rod is given its fundamental, then the wave-length of the sound in the rod is (§ 101) equal to $2L$, where L is the length of the rod. Hence for the material of which the rod is composed we have $V = 2nL$, while for the gas in the tube $v = 2nx$. Therefore

$$V/v = L/x$$

Thus, by measuring the ratio of the length of the rod to the distance between two loops, we can calculate the ratio of the velocities of sound in the material of the rod and in the gas. If we know the frequency n of the rod, the velocity of sound in the gas can be calculated, so that by filling the tube with various gases we can obtain the velocity of sound in these gases. Without knowing n we can, by simply comparing the values of the wave-length, obtain the ratio of the velocities in the different gases. For if v and v' are the velocities of sound in two gases and l and l' the wave-lengths corresponding to the tone of frequency n given by the rod, we have $v = nl$, and $v' = nl'$, so that

$$\frac{v}{v'} = \frac{l}{l'} = \frac{x}{x'}$$

where x and x' are the distances between consecutive loops as given by the Kundt's tube.

104. Maintenance of Vibrations.—When a tuning-fork is struck and thus set into vibration a certain amount of energy is communicated to the fork by the blow. As the fork continues to vibrate this energy is gradually dissipated, partly in the form of the sound-waves produced in the surrounding air and partly in friction between the prongs of the fork and the air, this dissipation of the energy being accompanied by a gradual decrease in the amplitude of the vibrations. If the fork is provided with a resonator the amplitude of the sound-waves is much increased, and hence the rate at which the energy of the fork is dissipated is increased.

Logarithmic decrement.

The decrease in the amplitude of the vibrations of a body which, after being set in vibration, is left to itself is said to be due to *damping*, so that the tuning-fork provided with a resonator is more damped than it is when no resonator is used. The extent of the damping is measured by the amplitudes of successive vibrations. In general this ratio is constant for a given set of conditions of the vibrating system. Thus if A_1 and A_2 are the amplitudes of two successive vibrations,

$$A_1/A_2 = \text{constant},$$

and hence

$$\log A_1 - \log A_2 = \text{constant}.$$

Thus the *difference* in the logarithms of the amplitudes of successive vibrations measures the damping, and this quantity is called the *logarithmic decrement.*

The amount of the damping of a body plays an important part in determining the amplitude of the forced vibrations set up when a periodic force acts on the body. Thus if the damping is small the amplitude of the forced vibrations are small unless the period of the force agrees almost exactly with the natural period of the body. That is, the resonance is very well marked. If, however, the damping of the body is great the amplitude of the forced vibrations is not very great, but this amplitude does not decrease very rapidly as the period of the force is changed from the natural period of the body, *i.e.* the resonance is not very well marked.

Maintenance of vibrations in an organ pipe.

If it is desired to maintain the vibrations of a body at a constant amplitude, then a continuous supply of energy must be provided in order to make up for the losses. In the case of an organ pipe, this supply of energy is provided by the stream of air from the bellows. The mouth of an ordinary flute-organ pipe is shown in Fig. 176. The wind is supplied through a tube A and escapes through a narrow slot. Opposite this slot is a bevelled lip B. If we suppose the pipe is sounding, the mouth will be a loop, and at a given instant let the air

be travelling outwards, as shown by the arrow C in (*a*). As a result the air-stream from E will be directed outwards, as shown by the arrow D. Half a period later the direction of motion of the air in the pipe will be reversed, as shown by the arrow C in (*b*), and as a result the air-stream escaping from E will now be directed into the pipe. Thus a series of puffs of air will occur in the pipe which are exactly timed so as to assist the motion of the air due to the vibrations, and thus the vibrations are maintained. When the stream of air is first started no vibrations are existing in the pipe, but the starting of the air-blast will cause a condensation to move up the pipe, and this will be reflected from the other end, and when it reaches the mouth will deflect the air-stream and thus the vibrations will be started, the amplitude gradually growing till the rate at which energy is being lost due to the sound produced is equal to the rate at which energy is being supplied by the air-blast. It will be observed that here we are dealing with vibrations caused by a periodic force, but the peculiarity is that the *period* of the force is settled by the natural period of the vibrations which are produced. A clock or watch is an example of a similar state of affairs.

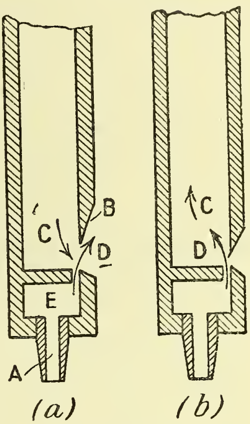


FIG. 176.

Another example of the above is what is called a singing flame, where the energy required to maintain the vibrations is supplied by a periodic supply of heat.

Singing flame.

If a jet of hydrogen gas be placed within a vertical tube open at both ends, then in general a loud note will be produced, which will continue as long as the gas jet remains alight. The same phenomenon is exhibited by burning jets of other combustible gases, but to a less marked degree. If the flame is observed by means of a rotating mirror, similar to that used in connection with manometric flames, it will be seen that the flame is in vibration. By using the stroboscopic method of observing the flame, Töpler was able to show that in many cases at one time during each vibration the flame retires inside the jet through which the gas is supplied. It is also found that the length of the gas supply-tube bears an important part in the phenomenon. If the supply-tube near the jet is lightly plugged with cotton-wool the gas flame, although it appears just as usual, is incapable of producing vibrations, while the notes which can be obtained with any given flame depend on the length of the supply-tube and on the nature of

the gas. These observations indicate that stationary waves are set up in the supply-tube. The effect of these vibrations in the supply-tube is that the emission of the gas, instead of being uniform, is intermittent, so that the size of the flame, and hence also the supply of heat to the air contained in the tube which surrounds the flame, is intermittent. Now when a column of air is in vibration and heat is supplied to the air at the moment of greatest condensation, this supply of heat will increase the force with which gas tends to expand, *i.e.* to regain its normal condition of pressure. The effect of this will be similar to that produced when a pendulum is struck a blow at the end of its swing tending to drive it back towards its position of rest, namely, it will tend to increase the amplitude of the vibrations. If, on the other hand, the supply of heat takes place when the air is at its greatest rarefaction, this will tend to resist the return of the air to its condition of rest, and will therefore tend to check the vibrations. Just as in the case of the pendulum, if it is struck a blow tending to check its motion as it is passing through its position of rest, the amplitude will decrease. Hence, if the periodic increase in the size of the flame always occurs at the instant when the air, in that portion of the tube near the flame, is, owing to the natural vibrations of the column of air in the tube, at its maximum condensation, the amplitude of the vibrations will be increased or at any rate maintained. If, however, the increase in size of the flame occurs sometimes at the instant of maximum condensation and sometimes at that of maximum rarefaction, that is, if the natural periods of the column of gas in the supply-tube and of the column of air in the tube are not commensurate, the heat will sometimes assist the vibrations and sometimes oppose. Hence, under these conditions, the vibrations of the air in the tube will not on the whole be maintained by the heat, and so will die out. It will thus be evident why it is necessary that the length of the supply-tube and the position of the flame should bear definite relations to the length of the tube in order that a sound may be produced. When a plug of cotton-wool is placed in the supply-tube vibrations can no longer take place in the gas contained in the tube, and so the *variations* in the size of the flame, which are necessary if the vibrations in the air column are to be kept up, are not produced.

CHAPTER VII

AUDITION

105. The Ear. Limits of Audition.—The human ear consists of an outer passage which is closed at its inner end by a membrane called the drum or tympanic membrane. The air waves travel down this passage, and striking the drum set it into forced vibrations. The centre of the drum is attached to a small hammer-shaped bone called the malleus, which is attached to a second small bone called the anvil or incus. This latter is attached to a stirrup-shaped bone, the stapes, and the part of this bone corresponding to the foot-plate of the stirrup is attached to a membrane which closes the inner part of the ear where the nerves terminate which convey the sense of sound to the brain. The inner ear consists of two parts, the three semicircular canals which appear to be used as a means of determining level and are not connected with audition, and a spiral body, the cochlea. It is in the cochlea that the auditory nerve terminates in a large number of filaments, called Corti's fibres. Thus sound-waves in the air first set the drum in vibration, and these vibrations are communicated by the auditory ossicles to the liquid which fills the cochlea, and it is the motion of this liquid which affects the nerve terminals.

The ear is not capable of detecting as sound the presence of air-waves of all frequencies, and it is only when the frequency of such waves falls between certain limits that the ear is able to distinguish their presence, and we experience the sensation we call sound. Neither of these limits is, however, well defined. Helmholtz concluded from his experiments that the lowest frequency which causes the sensation of a *musical tone* is about thirty vibrations per second. In forming any such estimate, it is very difficult to obtain a tone in which we may be quite certain no overtones are present; for, if they are present, what is actually heard may be the overtones and not the fundamental.

The upper limit of audibility is even more uncertain, for not only does it vary very much with the observer, but there is the added difficulty that it is very hard to determine the frequency of notes of very high pitch. The upper limit of audibility for normal ears appears to be somewhere between 10,000 and 20,000 vibrations per second.

Estimates of *pitch* cannot, however, be made above a frequency of about 4000.

A closely related subject is the amplitude of the sound-waves in air necessary for audition. As a result of some experiments on the distance to which a whistle could be heard when a measured power was employed in maintaining the sound, Rayleigh came to the conclusion that in favourable circumstances the ear is able to detect a sound, if the amplitude of the sound-wave exceeds 10^{-8} cm.

The direction from which a sound comes can be judged with considerable accuracy, and although the exact method by which we are able to make this estimate of direction is rather uncertain, there is no doubt that we are very largely guided by the effect of the sound on the two ears; probably the slight difference of the intensity with which the sound reaches the two ears is at the base of all such judgments when the sound is of high pitch, while for sounds of low pitch the difference in phase with which the waves reach the two ears is used.

106. Concord and Discord.—When two notes are sounded together some combinations produce a pleasing effect, when the interval is said to be *concordant* or *consonant*, and some a markedly displeasing effect, when the interval is said to be *discordant* or *dissonant*. It is found that the smaller the two numbers which represent the interval the greater is the concord. Thus unison, or 1/1, is the most concordant, then the octave 1/2, next the fifth 2/3, and so on, while an interval of a tone 9/8 is quite discordant. According to Helmholtz's theory of audition, each of the fibres of Corti has its own natural period, and when a note of any given period strikes the ear and reaches the cochlea, the fibres which agree in frequency with that of the note are set into vibration. If the note consists of a simple tone, that is if it contains waves of only one frequency, then only one, or at any rate only a few, of the fibres will respond.

When two tones of nearly the same pitch are sounded simultaneously, then some of the fibres will respond to both tones, but the vibrations set up will, on account of the production of beats, be intermittent in character. If this intermittence is sufficiently slow, the beats will be heard as separate maxima of sound. If, however, the beats are rapid, so that the fibres have not time to completely come to rest, or at any rate if there is not time for the nerve to recover completely between the stimuli, the effect will be noticed as a roughness or discord. When, however, the interval between the tones is so great that none of the fibres are set in vibration by *both* tones, then the sense of roughness will vanish.

If we further suppose that each thread may be capable of vibrating in more than one way, say the overtones are the harmonics of the funda-

mental, so that any given fibre would respond to a tone, and its octave, twelfth, &c., we can understand how it is that the interval of the octave is so consonant, and it would further explain why a tone and another, the frequency of which is nearly the same as the octave of the first, produce discord.

Starting with two tones of the same frequency, and then gradually increasing the frequency of one of them, the number of beats produced gradually increases. Very slow beats are not very unpleasant, but as the frequency of the beats increases so does the unpleasantness, till for a certain number of beats per second the discord is a maximum. If the number of beats is still further increased, the unpleasant sensation gradually decreases. The phenomenon is similar to that which occurs in the case of vision. If the intensity of a light alters, that is, the light flickers, the unpleasant sensation produced first increases as the frequency of the flicker increases, reaches a maximum and then decreases, and if the frequency is more than twenty per second all sense of flicker is lost, and the illumination appears continuous. The frequencies of the beats for which the discord is a maximum, and for which the sensation becomes continuous, vary with the absolute frequencies of the tones which give the beats.

When considering the accord or discord produced by compound tones, such as ordinarily occur in music, the presence of the upper partials must be taken into account, for although the fundamentals may be in accord, the upper partials of the notes may produce discord.

Of course the effects of the upper partials, in giving dissonance for any given musical interval, will depend on the strength of the partials and on the relation which their pitch bears to that of the fundamental, so that the source of two musical notes has to be taken into account when considering consonance. For simplicity we shall, however, suppose that the notes are such that the partials consist of the first seven harmonics of the fundamental. Hence the frequencies of the fundamental and the overtones are represented by the numbers 1, 2, 3, 4, 5, 6, 7, 8.

In each case we will assume the tonic to have a frequency of 256, and will then examine the frequencies of the partials of this tonic and of notes which together with it produce some of the commoner musical intervals.

In the first place, if the interval is the octave, so that the frequencies of the two fundamentals are 256 and 512, the odd overtones of the lower note agree in pitch with the frequencies of the fundamental and overtones of the higher note, so that the effect of the second note is to strengthen some of the partials of the lower note, and does not introduce any new element to produce discord.

The following table contains the frequencies of the partials of the

notes for some other intervals, the tonic having, as before, a frequency of 256 :—

	Tonic.	Fifth.	Fourth.	Major Sixth.	Major Third.	Minor Third.
Fundamental	256	384	341	427	320	307
1 overtone	512	768	682	854	640	614
2 ,,	768	1152	1023	1281	960	921
3 ,,	1024	1536	1364	1708	1280	1228
4 ,,	1280	1920	1705	2135	1600	1535
5 ,,	1536	2304	2046	2562	1920	1842
6 ,,	1792	2688	2387	...	2240	2149
7 ,,	2048	...	...	...	2560	2456
8 ,,	2304	...	...	...	...	...

In the case of the fifth, it will be observed that the difference in frequency between the fundamentals is 128, while this number also expresses the smallest difference which occurs between any of the partials. In this case the beats are too rapid to produce any discord.

In the fourth, the smallest difference in frequency is 83. There is, apparently, a difference of 2 between the seventh overtone of the fundamental and the fifth overtone of the higher note, but this is because for simplicity we have taken the frequency of this note as 341, when it ought to be 341·3.

In the major sixth, the smallest difference is 84.

In the major third, the smallest difference is 64.

In the minor third, the smallest difference is 50.

Now experiment shows that when the frequency of the lower note is 256, the maximum discord corresponds to about 20 beats per second, while most of the roughness vanishes when the frequency of the beats becomes 50.

Now of the intervals considered above the most consonant is the fifth, and the consonance decreases as we pass to the minor third. This decrease in the consonance is accompanied by a decrease in the smallest difference in the frequency of the partials of the two notes, so that in the case of the minor third we are approaching the limit (50 beats per second) at which discord begins.

Next let us take a case where, although the difference between the frequencies of the fundamentals is greater than in several of the cases above, yet the consonance is not so good, and see whether we can account for this dissonance by the production of beats between the partials. For this purpose we may take the notes *g*♭ and *g*♯ and a slightly mistuned fifth.

	c	$g\flat$	$g\sharp$	Mistuned Fifth.
Fundamental	256	369	400	380
1 overtone	512	738	800	760
2 ,,	768	1107	1200	1140
3 ,,	1024	1476	1600	1520
4 ,,	1280	1845	2000	1900
5 ,,	1536	2214	2400	2280
6 ,,	1792	...	...	...
7 ,,	2048	...	...	...
8 ,,	2304	...	...	...

In the case of c and $g\flat$, the difference between the fundamentals is 113, and so these tones will not produce discord. The second overtone of c and the first overtone of $g\flat$, however, differ by 30, and are therefore dissonant, and it is to the beats produced by these that the dissonance of the interval is due. In the case of c and $g\sharp$, the second overtone of c and the first overtone of $g\sharp$ differ in frequency by 32, while the seventh overtone of c and the fourth overtone of $g\sharp$ differ by 48, and the dissonance of the interval is thus accounted for. In the untrue fifth there will be eight beats per second between the second overtone of the lower note and the first overtone of the higher, sixteen per second between the fifth overtone of the lower and the third overtone of the higher, and twenty-four between the eighth overtone of the lower and the fifth of the higher. Hence it will be seen why it is that an untrue fifth is dissonant, and how the ear is able to detect want of correctness in such an interval.

Timbre.

Another influence of the overtones which ordinarily accompany a fundamental is to give quality or *timbre* to a note. Thus the note given by a tuning-fork, which is almost devoid of all overtones, is of a dull character, while the note given by the violin or by a cornet is composed of a fundamental accompanied by comparatively strong overtones, and it is to these overtones that the character of the note is due.

The presence of the various overtones in the note given by a musical instrument can easily be detected by having a resonator tuned to the frequency of the overtone to be looked for. If this overtone is present the resonator will be set into vibration. Thus by having a series of resonators of different pitch, or one of which the pitch can be altered, the frequencies of the overtones of a musical note can be picked out, that is, the note can be analysed into its component tones.

107. Production of Vocal Sounds.—The actual organ concerned in the production of the vibrations that constitute a vocal sound are two membranes, called the vocal cords. These membranes are stretched between a series of cartilaginous structures, to which are attached a series

of muscles, by means of which the tension of the membranes can be altered. The vocal cords are stretched across the opening of the trachea, which is a tube leading to the lungs, and it is to the vibrations caused in the cords when air is forced between them that vocal sounds are due. The vocal cords in men are thicker than in women and children, so that they vibrate more slowly, and hence produces lower notes. The sounds produced by the vocal cords are much modified by the effect of the mouth, which acts as a resonator of variable shape and volume.

HARMONICS CORRESPONDING TO THE VOWEL A

	Fundamental.	2	3	4	5	6	7	8	9	10
Frequency . .	113	226	339	452	565	678	791	904	1017	1130
Amplitude . .	9·9	15·4	2·5	0·8	11·5	16·1	3·3	8·1	21·2	7·9
Frequency . .	144	288	432	576	720	864	1008	1152	1296	1440
Amplitude . .	32·9	4·3	1·9	8·5	14·8	2·1	8·8	13·5	7·2	1·3
Frequency . .	171	342	513	684	855	1026	1197	1368	1539	1710
Amplitude . .	5·6	4·8	2·9	17·5	5·0	33·1	19·9	2·1	6·1	1·9
Frequency . .	226	452	678	904	1130	1356	1582	1808	2034	2260
Amplitude . .	9·2	3·6	34·6	8·0	31·0	2·6	6·2	1·4	0·8	1·2
Frequency . .	288	576	864	1152	1440	1728	2018	2304	...	...
Amplitude . .	5·4	18·2	8·0	57·3	3·0	1·3	3·2	3·6	...	...
Frequency . .	362	724	1086	1448	1810	2172	2534	2896	...	...
Amplitude . .	9·5	14·3	42·9	16·5	9·5	2·5	3·5	1·3	...	...
Frequency . .	426	852	1278	1704	2130	...	...	...	...	...
Amplitude . .	5·7	16·5	64·9	6·0	6·9	...	...	...	...	...
Frequency . .	576	1152	1728	2304	2880	...	...	...	...	...
Amplitude . .	21·4	68·0	4·0	5·6	1·0	...	...	...	...	...

Articulate speech is composed of a number of characteristic sounds called vowels, of which there exist almost an infinite number of different kinds, the characteristic being that a vowel sound can be indefinitely sustained without losing its characteristic. In addition to the vowel sounds, there are other sounds called consonants, which are not persistent sounds, being practically only different ways of commencing and ending a vowel sound.

The question as to what it is that gives its character to a vowel sound is a subject about which there has been, and is still, much difference of opinion.

This question has been investigated by Bevier by reproducing on an enlarged scale the form of the groove which constitutes the record in a

phonograph, and then analysing the resulting curve according to Fourier's theorem; that is, he determined the amplitudes of the S.H.Ms. which when compounded together in the manner described on page 222 would produce this curve. A copy of a portion of the curve corresponding to the vowel *ah* is shown in Fig. 177. The amplitudes of the fundamental, that is, the note on which the vowel was sung, and of the harmonics obtained by the Fourier analysis of the curves, are given in the table on the previous page.

It will be noticed that the *order* of the harmonic which has the maximum amplitude varies with the pitch of the fundamental, but the *pitch* of the harmonic which has the greatest amplitude always lies between 1000 and 1300 vibrations per second, the maximum effect

Fig. 177.

occurring when one of the harmonics has a frequency of about 1150. Harmonics having frequencies between 575 and 800 for men, and between 675 and 900 for women, are also prominent, but the amplitude is less than in the case of the higher maximum. The fundamental varies greatly in amplitude. In the same way Bevier finds that *i* (as in pit) is characterised by the following: (1) A single harmonic, of frequency about 1850, is very strong; (2) harmonics having a frequency about 575 are fairly strong, the harmonics between this and (1) being very weak. The amplitude of the fundamental varies considerably according to its pitch. If the frequency is 200, or above 500, the amplitude is large; but if the frequency lies between 275 and about 450, the amplitude is small. Similar results have been obtained for the other vowels.

BOOK IV

LIGHT

CHAPTER I

REFLECTION

108. Rectilinear Propagation of Light, Shadows, Pinhole Camera, Parallax. Images.—The term light is used both in a subjective sense, when our eye is subjected to certain influences which produce the sensation which we call light, and also in the objective sense, when referring to the physical cause of the above sensation. As we shall see later, it is convenient to study at the same time the properties of other kinds of radiation which, although they physically resemble light, yet do not produce in the eye the sensation of light. For the present we shall postpone the consideration of the nature of light, contenting ourselves with the statement that light is a periodic motion which is propagated in the form of waves in some medium which, since light comes to us through interstellar space, is obviously not matter.

The most striking property of light is that in an isotropic medium it is propagated in straight lines. A line drawn so as to represent the direction in which the light is propagated is called a ray, while a number of rays are often referred to as a beam or pencil of light.

Shadows.

When an opaque body is placed in the path of a beam of light, consisting of a number of rays emanating from a point, such as would be produced by a very small luminous source, then the light will be cut off by the obstacle from a region behind the obstacle enclosed by a series of straight lines drawn from the source to touch the boundary of the obstacle. This region is called the *shadow* of the obstacle, and if a screen be placed behind the obstacle the shadow will have a sharp boundary. If the source of light has an appreciable magnitude, however, we do not get a simple shadow of uniform blackness with a sharp outline. Let AB (Fig. 178) be a luminous object, say the sun, and CD the body that casts the shadow, say the moon. Then if we consider a point, A, of the luminous body, the shadow cast by this point on a screen at EF would be at HK. In the same way the shadow cast by the point B would be GI. All intermediate points would cast shadows situated between G and K. It will thus be seen that HI will be the only part of the screen which is completely in shadow, *i.e.* screened from the whole of the luminous object. This part of the shadow is therefore called

the *umbra*. The rest of the shadow is not completely dark, but gets darker and darker from the outside to the edge of the umbra. This part of the shadow is called the *penumbra*. In the case of the moon and earth, it is only when the earth enters within the cone CMD that a total eclipse takes place; when it enters within the penumbra the eclipse is only partial, since from any point within the penumbra straight lines can be

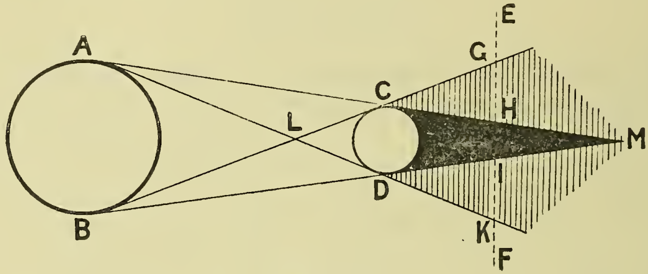


FIG. 178.

drawn touching the object, which will intersect the source of light, and so part of the source will be visible from any such point.

The working of the pin-hole camera depends on the rectilinear propagation of light. If a small hole is made in an opaque screen, and a luminous object is placed on one side and a white screen on the other, an inverted image of the luminous object will be formed on the screen. Each luminous point of the object, such as A and B (Fig. 179), will form a small round patch

The pin-hole camera.

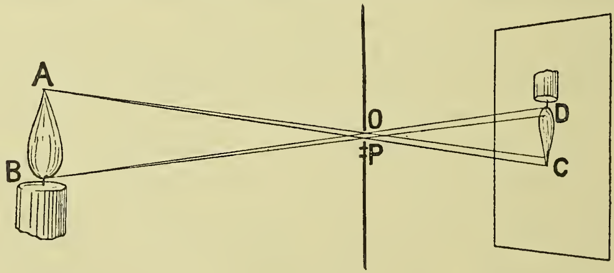


FIG. 179.

of light on the screen; and if the hole is so small that these patches of light do not overlap very much they will build up an image of the object, which, as is shown in the figure, is inverted. It is important to note that the image will be formed whatever may be the relative distance of the object and screen from the pin-hole, so that in this particular we

have an important difference between the image formed in this way and that produced by a lens or mirror (§§ 110, 113). If a second pin-hole were made near the first, say at P, a second image would be produced, which would partly overlap the first image. In the same way if a number of holes were made surrounding O, instead of a definite image we should simply have a blur produced by the partial superposition of all the images. This explains why it is that it is only when the pin-hole is small that any sharp image is obtained, for a large hole is the equivalent of a number of pin-holes close together.

Since the triangles AOB and DOC are similar, the size of the image is to the size of the object as the distance of the image from the pin-hole is to the distance of the object from the pin-hole.

Parallax.

Suppose we have two objects, A and B (Fig. 180), and that an observer is placed at C, then A will appear to be to the right of B. If, however, the observer moves across to D, A will now appear to the left of B. Hence the relative positions of A and B will appear to change as the observer moves unless they are at equal distances away. This phenomenon is called *parallax*, and it enables an observer to tell whether A or B is the nearer, even when by observing from one position, such as C, it is impossible to distinguish their relative distances. It is obvious that the object which appears to be displaced in the *opposite* direction to that in which the observer moves is the *nearer*.

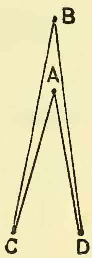


Fig. 180.

Image.

In certain cases the directions of the rays of light proceeding from a body do not travel throughout their course from the body to the eye in the same straight line. If, owing to reflection or refraction or both together, the rays of light which leave a certain point A of a luminous body and enter the eye, when they enter the eye proceed as if they came from some different point B, then the luminous point A will appear to be situated at B, and B is called the *image* of A. Thus if after reflection, refraction, or the like, the direction of the rays of light which emanate from a luminous point A pass through a second point B, then B is said to be the image of A. There are two kinds of images. In the case of the first kind, called *real images*, the rays of light which leave the point A after reflection or refraction actually pass through the image point B. In the second kind, called *virtual images*, the rays after reflection or refraction do not actually pass through the image B; they do, however, proceed as if they came from B. Thus in the case of a virtual image, if the directions of the rays after leaving the last reflecting or refracting surface are produced *backwards* they would intersect at the image.

Real images can be received on a screen and viewed in that way. Not so virtual images, these can only be perceived when the rays enter the eye directly, the images appearing suspended in space. They can, however, be located by the parallax method described above.

109. Reflection of Light at Plane Surfaces. The Sextant. —The fact that bodies which are not themselves luminous are, when illuminated, visible in all directions, shows that they must be capable of reflecting light in all directions, for it is by these reflected rays that we are able to see the body. Such rays, which are reflected from a body in all directions, and which often differ in many ways, such as colour, from the incident light, are said to have undergone diffused reflection.

When a beam of light is incident on a well-polished mirror, it undergoes diffused reflection to only a very small degree, the greater part of the light being reflected in a single direction. This light is said to have undergone regular reflection, and we now proceed to consider the laws that govern regular reflection.

The laws of reflection.

The point where a ray of light strikes a mirror is called the point of incidence. If through the point of incidence a line, called the normal, is drawn at right angles to the reflecting surface, the angle the incident ray makes with this line is called the angle of incidence, while the angle the reflected ray makes with the normal is called the angle of reflection. The phenomena of regular reflection may then be summed up in the following two laws :—

1. *The incident ray, the normal to the reflecting surface at the point of incidence, and the reflected ray, are all in the same plane.*

2. *The angle of reflection is equal to the angle of incidence.*

The truth of these laws can be proved by observing the position of the image formed by reflection in a plane mirror. Thus suppose an object P (Fig. 181) is placed before a plane mirror AB, then a single image of P will be observed. This image is virtual and is situated at the point P′ wherever the eye of the observer is placed. Thus if the eye is at Q_1 the light from P, which after reflection enters the eye, travels along the line M_1Q_1, and the direction of this line when produced backwards passes through P_1. Similarly if the eye is at Q_2, and so on. The position of P′ can be fixed by making use of the parallax method, and it is found that the line joining the object P to the image P′ is at right angles to the mirror, and that P′

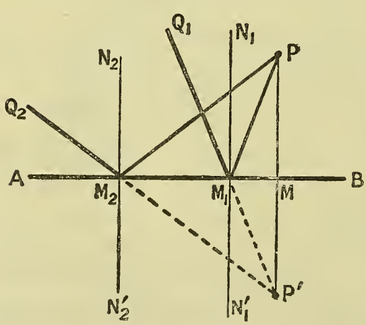


FIG. 181.

is as far *behind* the mirror as P is in front. Then since $PM = P'M$ and PP' is at right angles to AB, the triangles PM_1M and $P'M_1M$ are similar. Hence the angle PM_1M is equal to the angle $P'M_1M$. Thus if $N_1M_1N'_1$ is the normal at M_1, the angle PM_1N_1 is equal to the angle $P'M_1N'_1$. But, since Q_1M_1P' is a straight line, the angle $Q_1M_1N_1$ is equal to the angle $P'M_1N'_1$, and therefore the angle of reflection $Q_1M_1N_1$ is equal to the angle of incidence PM_1N_1. Thus the second law is verified. In the same way experiment shows that the points P, M_1, and P′ all lie in a plane which also contains the normal N_1M_1, and hence the first law is true.

Rotation of a plane mirror.

If a ray of light is reflected from a plane mirror and the mirror is rotated about an axis perpendicular to the plane of incidence, the *reflected* ray will be rotated through an angle *twice* as great as that through which the mirror turns. We may prove this proposition in the following manner, in which an important proposition as to the position of the image is first proved.

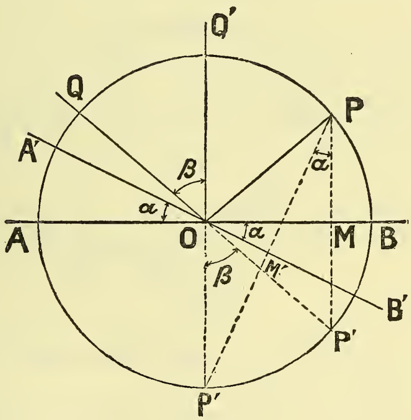


FIG. 182.

Let O (Fig. 182) be the point about which the mirror turns, then if a circle be described with O as centre and OP as radius, we shall show that whatever the position of the mirror, the image of P will lie on this circle. Let AB be a position of the mirror, then if P′ is the image of P, we have that PP′ is perpendicular to AB, and that $PM = P'M$. Hence in the two triangles OPM and OP′M, the two sides OM, MP of the one are equal to the two sides OM, MP′ of the other, and the included angles are equal, each being a right angle. Hence the triangles are equal in all respects, and OP is equal to OP′, hence P′ is on the circle described with O as centre and OP as radius. If the mirror is rotated about the point O through an

angle α to the position A′B′ and P″ is the image of P, then we can show just as before that P″O = OP, and hence P″ must be on the circle.

Since PM and PM′ are the normals to the mirror in the two positions, the angle M′PM or P″PP′ is equal to α, and by a well-known property of the circle, the angle P″OP′ subtended by the arc P″P′ at the centre is twice the angle P″PP′ subtended by the same arc at a point on the circumference. Hence the angle P″OP′ is 2α, so that when the mirror turns through an angle α, the line joining the axis of rotation to the image turns through an angle 2α. Now if PO is an incident ray, then OQ and OQ′ will be the reflected rays in the two positions of the mirror. Hence the angle QOQ′ included between the two reflected rays is 2α.

The deflection of a ray of light reflected from a mirror is often used to measure the rotation of the mirror, and hence that of some piece of apparatus, such as a galvanometer needle, to which the mirror is attached.

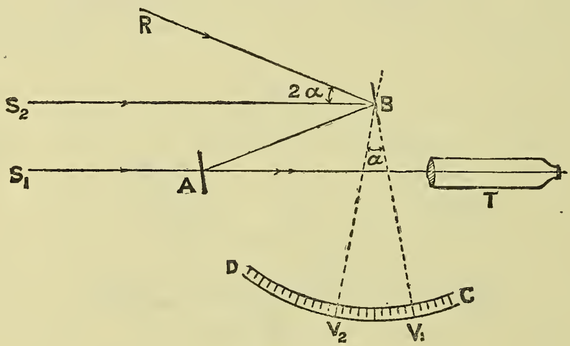


FIG. 183.

The ray of light forms a pointer which may be quite long and yet is without weight and does not bend.

The principle that the reflected ray is turned through twice the angle that the mirror is turned, is made use of in the sextant to measure the angle subtended at the observer by two distant objects, say the sun and the horizon. The sextant consists essentially of two mirrors, one of which, A (Fig. 183), is fixed relative to the instrument, while the other, B, is movable about an axis at right angles to the paper, the angle through which it is turned being read by means of a vernier, V, and a graduated circular arc DC. The upper part of the mirror A has the silver removed, so that it is transparent, and a telescope T is so placed that half the light that enters the object-glass comes *through* the upper part of the mirror A, and the rest is light which has been reflected in the lower part. Suppose S_1A is a ray of light coming from a distant object, which traverses A and enters the telescope, and thus helps to produce an image of the object.

Now the movable mirror B can be so turned that a ray S_2B, coming from the same object, after reflection in the mirrors B and A, also enters the telescope. This ray will help to produce a second image of the object, which, by rotating the mirror B, can be brought alongside the image formed by the direct rays. Next, keeping the telescope turned so that the direct rays S_1 still form an image, turn the mirror B till the rays RB, proceeding from some other object, enter the telescope and form an image alongside the first. Since merely reversing the direction of the rays of light will not alter their paths, we now see that while in the first position of the mirror a ray incident along AB is reflected along BS_2, in the second position a ray incident along AB is reflected along BR. Hence, if α is the angle through which the mirror has been turned, the angle RBS_2 is equal to 2α. Now the angle RBS_2 is the angle subtended at B by the two objects, and so, since α is obtained by reading the two positions of the vernier on the arc, this angle can at once be obtained. In order to save the necessity of doubling the reading of the scale, it is usual to number each half-degree as a whole degree, so that the reading gives directly the value of 2α.

The sextant.

If an object P (Fig. 184) is placed between two mirrors AO, BO which are inclined at an angle, then a series of images will be produced. Thus, if with O as centre and OP as radius we describe a circle, the image of P produced in AO will be at P′ on this circle, and the arc MP will equal the arc MP′. Similarly the image Q′ in the mirror BO will lie on the circle, and the arc Q′N will be equal to the arc PN. Now the image P′ will act as an object as far as the mirror BO is concerned, and its image will be at Q″, where the arc NP′ is equal to the arc NQ″. The rays of light which proceed as if they came from the image Q″ will have undergone *two* reflections, the first one in the mirror AO and the second in the mirror BO. Thus a ray of light which enters an eye at E and proceeds as if it came from the image Q″ would travel along the path PCDE. Similarly the image P″ will be formed by rays which have been reflected first in the mirror BO and then in the mirror AO. Finally the image Q″ will act as an object as far as the mirror AO is concerned, and will produce an image at P‴, the rays forming this image having been reflected three times, first in AO, then

Images formed by two inclined mirrors.

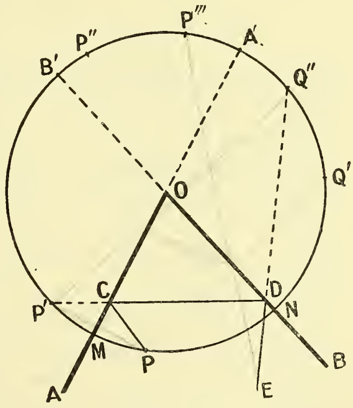


FIG. 184.

in BO, and finally in AO. The images P″ and P‴ can produce no more images since they both lie *behind* the planes of both mirrors, and it is obvious that for an object to be reflected in a mirror it must lie in front of the plane containing the mirror.

The number of images produced depends on the inclination of the mirrors, and it is easily seen that as the angle between the mirrors decreases the number of images possible increases. If we imagine the points M and N of the mirror to be fixed, as the angle between the mirrors decreases the point of intersection, O, of the mirrors gets further and further away, and hence the radius of the circle through P increases. When the mirrors are parallel, the point O is at infinity and the portion of the circle near the mirrors consists of a straight line through P perpendicular to the two mirrors, and all the images must lie in this line. Further, the number of images is now infinite. It can easily be shown

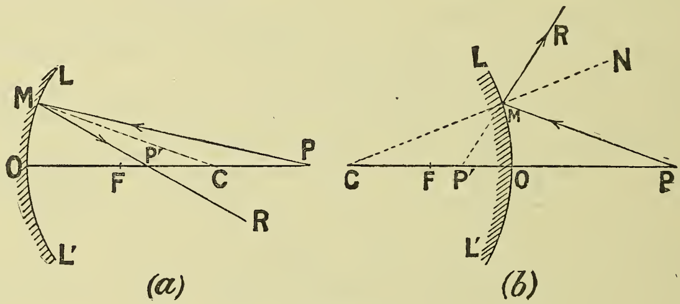


FIG. 185.

that the distance between successive images on either side is alternately equal to twice the distance of the object P from either of the two mirrors.

110. Reflection of Light at a Spherical Surface.—A polished spherical surface, where the reflection takes place from the inside surface, is called a *concave* spherical mirror, while one where the reflection takes place from the outside surface is called a *convex* spherical mirror.

A straight line passing through the mid-point of the mirror and the centre of the spherical surface of which the mirror surface forms a part, is called the *axis* of the mirror. The centre of the spherical surface is called the centre of curvature of the mirror, and the radius of the sphere is called the radius of curvature.

In the case of either type of mirror, let LOL′ (Fig. 185) represent a section of the mirror, OC being the axis and C the centre of curvature.

Suppose that a luminous point P is placed on the axis of the mirror, then the ray of light from P incident along the axis will be reflected straight back, for the line PCO is normal to the mirror at O. If we join the point of incidence M, where another ray PM strikes the mirror, to the centre C, then the line CM in Fig. 185 (*a*), or CM produced in Fig. 185 (*b*), will be the normal to the mirror at M; and hence if we make the angle RMC equal to the angle PMC, or, in the case of the convex mirror, the angle RMN equal to the angle PMN, the line MR will be the path of the reflected ray. Let MR, or in the case of a convex mirror MR produced backwards, cut the axis at P′, then P′ will be the point of intersection of the two rays incident at O and M respectively after reflection.

Taking the case of the concave mirror first, we have, since in the triangle P′MP the line CM bisects the angle P′MP, the following relation (*Euclid*, vi. 3)—

$$\frac{CP}{CP'}=\frac{PM}{P'M}$$

If the angle LCL′ subtended at the centre by the mirror, called the *aperture* of the mirror, is small, then PM is very nearly equal to PO, and P′M very nearly equal to P′O, and in these circumstances

$$\frac{CP}{CP'}=\frac{PO}{P'O} \quad . \quad . \quad . \quad (a)$$

Thus, if the mirror is of small aperture, the position of P′ does not depend on the position of the point of incidence M, but only on the distance PO and the radius of curvature OC of the mirror. Hence all the reflected rays will pass through P′, and P′ will be the image of P produced by reflection in the mirror. We shall for the present confine our attention to mirrors of such small aperture that the assumptions made above hold good, so that P′ will be the image of the point P formed by reflection in the mirror, and since the reflected rays actually pass through P′, the image is real.

We have now to make some convention as to the direction we shall call positive, and shall take all distances measured *from the mirror* in an *opposite* sense to that in which the incident light falls upon the mirror as *positive*, while all distances measured in the same sense as the incident light we shall take as *negative*.

Thus the distance OC, Fig. 185 (*a*), being measured in the opposite sense to the incident light, which proceeds from P to the mirror in the sense PO, is positive, while the distance OC, Fig. 185 (*b*), is negative. It will also be convenient to use single letters to represent some of the

distances which continually occur. We shall, therefore, in future indicate the radius of curvature OC of the mirror by r, the distance OP of the *object* from the mirror by u, and the distance OP′ of the *image* from the mirror by v.

Now $CP = OP - OC = u - r$; and $CP' = OC - OP' = r - v$; hence the equation (a) reduces to

$$\frac{u-r}{r-v}=\frac{u}{v}$$

or

$$uv - vr = ur - uv.$$

$$\therefore \quad 2uv = vr + ur,$$

and dividing all through by uvr, we get

$$\frac{2}{r}=\frac{1}{u}+\frac{1}{v} \quad . \quad . \quad . \quad . \quad (115)$$

This equation gives us the general relation between the distances of object and image from a concave mirror of small aperture in terms of the radius of curvature of the mirror.

Returning to the case of a convex mirror, Fig. 185 (b), it will be noticed that the reflected rays do not actually pass through the image P′, but only their directions, so that the image is virtual.

As in the case of the concave mirror, we have

$$\frac{CP}{CP'}=\frac{PM}{P'M}$$

and when the mirror is of small aperture, this reduces to

$$\frac{CP}{CP'}=\frac{PO}{P'O}$$

Now $CP = OC + OP = -r + u$, since OC is equal to $-r$, for the distance is measured in the negative direction, and OP is u; also $CP' = OC - OP' = -r + v$, both r and v being measured in the negative direction. Hence

$$\frac{u-r}{v-r}=\frac{u}{-v}$$

or

$$-uv + vr = uv - ur.$$

$$\therefore \quad 2uv = vr + ur.$$

$$\therefore \quad \frac{2}{r}=\frac{1}{u}+\frac{1}{v}$$

Thus we have the same equation as in the case of the concave mirror. In making any numerical application of this formula it must, however,

be carefully borne in mind that while for both classes of mirrors u is always positive, in the case of concave mirrors r is positive, while in the case of convex mirrors r is negative.

If we make the distance, u, of the object from the mirror larger and larger till the object is at an infinite distance, $1/u$ will become zero. Hence, in these circumstances,

$$v=\frac{r}{2} \quad . \quad . \quad . \quad . \quad . \quad (116)$$

and hence the image is formed at a point half-way between the mirror and the centre of curvature.

Since, when the object is at an infinite distance, all rays proceeding from it which strike the mirror may be considered as parallel, a pencil of parallel rays incident parallel to the axis is brought to a focus at a point on the axis at a distance equal to half the radius of curvature from the mirror. This point is called the *principal focus*, and the distance between it and the mirror is called the *focal length* of the mirror. In the case of a mirror the focal length (f) is equal to half the radius of curvature. Hence, in terms of the focal length, the formula for giving the position of the image becomes

$$\frac{1}{f}=\frac{1}{u}+\frac{1}{v} \quad . \quad . \quad . \quad . \quad (117)$$

Since the principal focus is half-way between the mirror and the centre of curvature, the sign of f is the same as that of r, that is, f is positive for concave mirrors and negative for convex mirrors.

To find the image of a small object on axis of a spherical mirror.

We have hitherto considered the image of a single luminous point, and now proceed to find the image of a small object placed on the axis of a mirror. Let PQ (Fig. 186) be such an object, then we may consider each point of the object as a luminous point and find its image, and all the images thus found will build up the image of the small object. The problem is most easily solved by a geometrical construction based on the results we have obtained above.

Consider the point P of the object. A ray incident along the line PCN, passing through the centre of curvature of the mirror, will meet the mirror normally at N and be reflected straight back along its path. A ray PM incident parallel to the axis of the mirror will, after reflection, either actually pass through the principal focus F (concave mirror), or its direction when produced back will pass through the focus F (convex mirror). These two rays, both proceeding from the point P, will therefore meet at P′, and this point will be the image of P. In the same way,

the image of Q can be found by the intersection of one ray passing through the centre of curvature, which will be reflected back on itself, with another taken parallel to the axis, which, after reflection, will pass through the principal focus. We thus obtain the images of the extreme points of the object, and may fill in the intervening part free-hand, since the image and object will be similar. It will be seen that for the positions shown the images are in both cases smaller than the object, and that the image in the concave mirror is inverted and real, while that in the convex mirror is erect and virtual.

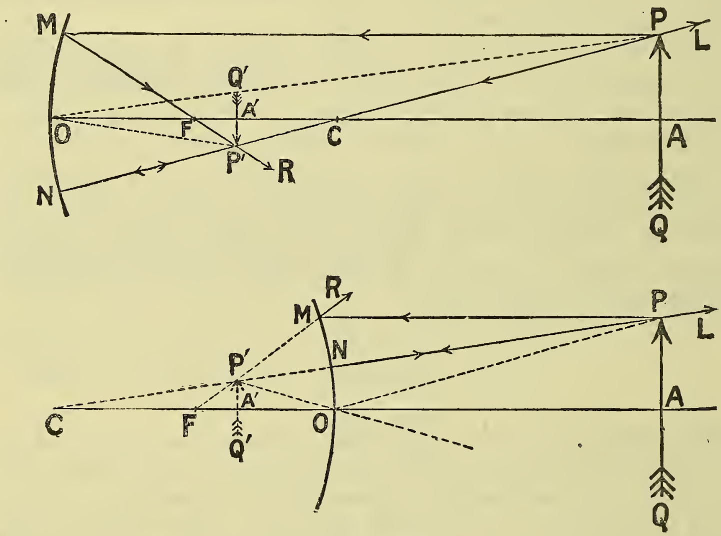


Fig. 186.

Relative size of image and object.

The relative sizes of the image and object can also be obtained from these figures, for a ray from P incident along PO will be reflected along OP′ to the image P′, and the angle of incidence POA must be equal to the angle of reflection AOP′. Hence the two triangles POA and P′OA′ are similar, and therefore

$$\frac{\text{size of object}}{\text{size of image}}=\frac{PA}{P'A'}=\frac{OA}{OA'}=\frac{u}{v} \quad . \quad . \quad . \quad (118)$$

Also, since the triangles PCA and P′CA′ are similar,

$$\frac{PA}{AC}=\frac{P'A'}{A'C}$$

Hence the ratio of the size of the object to that of the image is as the

ratio of their distances from the mirror, or as the ratio of their distances from the centre of curvature.

The construction used above may be employed to show very clearly how the position, size, &c., of the image changes as the distance of the object from the mirror is varied.

Draw any line PM (Fig. 187) parallel to the axis, and through the point M, where this line cuts the mirror, and the principal focus F draw the straight line R′MFR. If from any point in PM we draw a straight line through the centre of curvature, then where this line intersects the line RFR′ is the image of the given point. Thus the image of the point 1 is at 1′, and if perpendiculars are drawn to the axis, 1‴ is the image of 1″.

A study of the figure will show that in the case of a concave mirror, Fig. 187 (*a*), when the object is anywhere between infinity and the

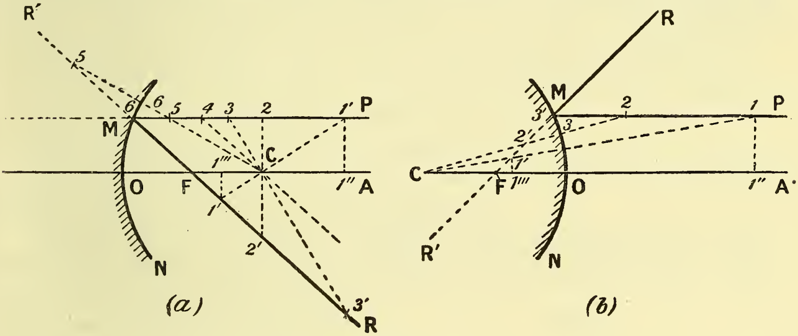


Fig. 187.

centre of curvature C the image is real, inverted, smaller than the object, and situated between the principal focus F and the centre of curvature C. If the object lies between the centre of curvature and the principal focus (point 3) the image is real, inverted, magnified, and situated between the centre of curvature and infinity. If the object lies between the principal focus and the mirror (point 5) the image is virtual, erect, magnified, and situated behind the mirror.

In the case of the convex mirror the image is always erect, virtual, smaller than the object, and situated between the principal focus and the mirror.

If the aperture of a concave spherical mirror is not small (say less than 20°), then by drawing a figure to scale it will be at once seen that, in the case say of a very distant point source of light, the rays which are reflected near the edge of the mirror intersect the axis at points nearer the mirror, and hence the point at which rays reflected near the

centre cut the axis. Hence in such a case all the reflected rays do not pass through a *single* point, and hence a blurred image is produced. This effect is referred to as *spherical aberration.* Similarly a point source placed at the principal focus only gives a beam of *parallel* reflected rays if the aperture of the mirror is small. If the aperture is large only the rays reflected from the central part of the mirror are parallel, the rays reflected from the peripheral parts of the mirror will converge towards a point on the axis, and after passing through this point will diverge. Thus when a concave mirror is used as a projector, say in a search-light, or motor head-lamp, the resultant beam will be partly parallel and partly divergent, so that the light will not be all concentrated in one direction. If the section of the mirror is a parabola, however, and the source of light is at the focus, *all* the reflected rays are parallel. Hence parabolic reflectors are always used in search-lights and the like.

CHAPTER II

REFRACTION

111. Refraction at a Plane Surface.—In general when a ray of light passes from one medium to another the direction of the ray changes at the surface of separation between the media. In addition to the light which penetrates the second medium, some of the light will be reflected at the surface of separation according to the laws considered in the previous chapter. In this chapter we shall in general neglect the reflected light and confine our attention to the light which penetrates the second medium, and which forms what is called the *refracted ray.*

The angle the refracted ray makes with the normal to the surface of separation is called the angle of refraction.

The laws governing refraction are as follows :—

(1) The refracted ray lies in the plane of incidence, and on the opposite side of the normal to the incident ray.

(2) The sine of the angle of refraction bears a constant ratio to the sine of the angle of incidence for all angles of incidence, the value of the ratio depending on the nature of the two media at the surface of separation between which the refraction takes place, and also on the nature of the incident light (Snell's law).

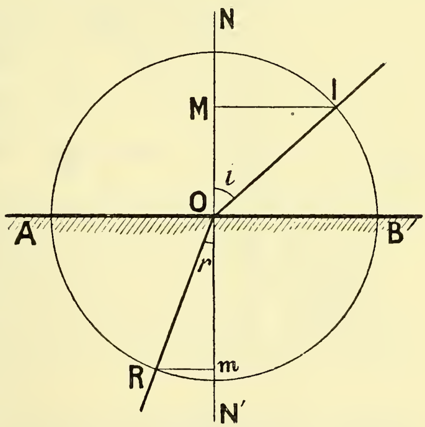


Fig. 188.

It will thus be seen that in the case of refraction the conditions are more complicated than in that of reflection, for while in the latter the *direction* of the reflected ray was independent of the nature of the reflecting surface, of the medium in which the light was travelling, and of the nature (colour) of the light, in the case of refraction the direction of the refracted ray depends on all these conditions. We shall for the present postpone the consideration of the influence of the nature of the light, assuming that

the light with which we are about to deal is the yellow light given out by a Bunsen flame when a bead of common salt (NaCl) is placed in the flame.

Suppose AB (Fig. 188) to be the surface of separation between two media, say air above and glass below, and that a ray of light travelling in the direction IO is incident at O. Let NON′ be the normal to the surface of separation at O, then in the case considered, in which the medium above AB is less dense than that below, the angle of refraction RON′, or r, will be less than the angle of incidence ION, or i. If the medium above AB had been denser than that below, then i would have been less than r.

If, with O as centre, we describe a circle of any radius, cutting the incident ray at I and the refracted ray at R, and from I and R draw perpendiculars to the normal, then

$$\sin i = \frac{IM}{IO}; \text{ and } \sin r = \frac{Rm}{RO}$$

or

$$\frac{\sin i}{\sin r} = \frac{IM}{Rm}$$

According to Snell's law the ratio $\sin i / \sin r$ is constant for all angles of incidence, and the value of this ratio for any pair of media is called the *refractive index* for these media, and is generally indicated by the Greek letter μ or the letter n.

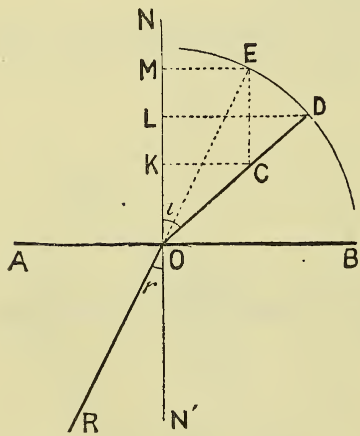


FIG. 189.

When the incident ray is perpendicular to the surface of separation i is zero, and hence $\sin i = 0$, so that $\sin r = 0$ and $r = 0$. Thus in this case the ray does not suffer refraction.

If we are given the refractive index between two media and the angle of incidence, it is easy, by a geometrical construction, to find the direction of the refracted ray. If AB (Fig. 189) is the surface of separation between the media, the denser being below, and DO is the direction of the incident ray, measure off from O along OD a distance OC equal to unity, and a distance OD, which, expressed in the same units, is equal to the refractive index n. With centre O and radius OD describe an arc of a circle. From C draw CE parallel to the normal ON, cutting the circle in E, join EO and produce to R, then OR is the refracted ray.

To prove this, draw EM, DL, and CK perpendiculars to ON.

Now $$\frac{DL}{OD} = \sin i, \text{ and } \frac{ME}{OE} = \sin r.$$

Hence $$\frac{\sin i}{\sin r} = \frac{DL}{OD} \cdot \frac{OE}{ME} = \frac{DL}{KC}$$

for OD and OE are equal, being radii of the circle, and ME is equal to KC. Now the triangles DOL and COK are similar. Hence

$$\frac{DL}{KC} = \frac{OD}{OC}$$

But by construction $$\frac{OD}{OC} = n$$

Hence $$\frac{\sin i}{\sin r} = n$$

and r is the angle of refraction, so that OR is the direction of the refracted ray.

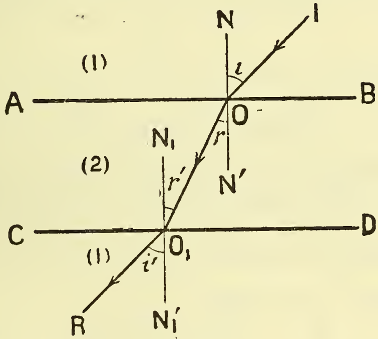


FIG. 190.

Refraction through a slab with parallel sides.

Suppose we have a slab of a denser medium enclosed by parallel sides AB and CD (Fig. 190), with a less dense medium on either side. Then it is found experimentally that if a ray of light is passed through the plate, the direction O_1R of the ray after leaving the plate is parallel to the incident direction IO, the only effect of the interposition of the plate being to displace the ray to one side.

We will call the less dense medium 1, and the medium composing the slab 2, and indicate the refractive index from medium 1 to medium 2 by ${}_1n_2$, and that from 2 to 1 by ${}_2n_1$. We have

$${}_1n_2 = \frac{\sin i}{\sin r} \text{ and } {}_2n_1 = \frac{\sin r'}{\sin i'}$$

Since the sides of the slab are parallel, and NON′ and $N_1O_1N_1'$ are normals these lines are parallel, and therefore the angle r is equal to the angle r'; also, since the rays IO and O_1R are found by experiment to be parallel, the angle i is equal to the angle i'.

Hence $$\frac{\sin i}{\sin r} = \frac{\sin i'}{\sin r'}$$

∴ $${}_1n_2 = \frac{1}{{}_2n_1}$$

or $${}_1n_2 \cdot {}_2n_1 = 1 \quad . \quad . \quad . \quad . \quad . \quad (119)$$

Thus we get that the refractive index from medium (1) into medium (2) is the reciprocal of the refractive index from medium (2) into medium (1).

By taking a number of slabs of media of different refrangibility, it can be shown, using a similar notation to that employed above, that

$${}_1n_2 \cdot {}_2n_3 \cdot {}_3n_4 \cdot {}_4n_5 \quad . \quad . \quad . \quad {}_nn_1 = 1 \quad . \quad . \quad . \quad (120)$$

This expression will be of use in solving problems on refractive index. Thus, given that the refractive index from air to glass is 1·5, and that from air to water is 1·33, find the refractive index from water to glass. In the first place,

$$n(air\ to\ glass) \times n(glass\ to\ air) = 1,$$

∴ $$n(glass\ to\ air) = \frac{1}{1 \cdot 5} = \cdot 67 \ldots$$

Also $$n(air\ to\ water) \times n(water\ to\ glass) \times n(glass\ to\ air) = 1,$$

∴ $$1 \cdot 33 \times n(water\ to\ glass) \times \cdot 67 = 1.$$

Hence $$n(water\ to\ glass) = \frac{1}{1 \cdot 33 \times \cdot 67} = 1 \cdot 13,$$

or the refractive index from water to glass is 1·13.

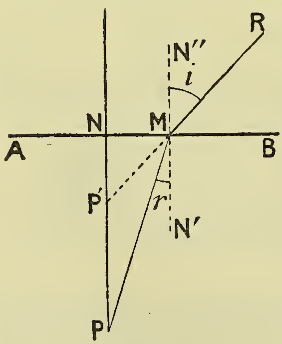


FIG. 191.

Apparent depth of water.

It is a familiar observation that water of such a depth that we can see the bottom always appears considerably shallower than it really is. This effect is due to refraction. Thus let P (Fig. 191) be a point on the bottom, and AB be the surface. A ray PM will at M be refracted along MR, and since it is proceeding from a more dense medium (water) to a less dense one (air), it is refracted away from the normal. Thus the ray MR in the air proceeds as if it came from a point P′. All the rays which could enter an eye placed at R would be refracted very near the point M, and the directions of all these rays would pass through a single point P′

which is the virtual image of P, and if the point M where the rays leave the surface is not very far from the foot N of the normal, so that the angle of incidence is small, the image P′ lies on the normal through P.

Since the angles NMP′ and P′MN′ together make up a right angle, the sine of one is equal to the cosine of the other. Hence

$$\sin RMN'' = \sin P'MN' = \cos NMP' = \frac{NM}{P'M}$$

Also since PN and N′MN″ are both normals to AB, they are parallel, and the angles PMN′ and NPM are equal. Hence

$$\sin PMN' = \sin NPM = \frac{NM}{PM}$$

But if n is the refractive index from air to water,

$$n = \frac{\sin RMN''}{\sin PMN'} = \frac{NM}{P'M} \cdot \frac{PM}{NM} = \frac{PM}{P'M}$$

If M is very near N, that is for nearly normal incidence, PM will be nearly equal to PN and P′M to P′N. Hence in this case

$$PN = \frac{PN}{n} \quad . \quad . \quad . \quad . \quad (121)$$

In the case of water the value of n is 1·33 or 4/3. Hence PN $= \frac{3}{4}$PN, and the water only appears as if it were three-quarters of its real depth.

The greater the angle of incidence of the rays which enter the eye, the shallower the water will appear, as can be easily proved by carefully drawing a diagram to scale.

The equation expressing Snell's law may be written

$$\sin r = \frac{\sin i}{n} \quad . \quad . \quad . \quad . \quad (a)$$

Total internal reflection.

If we are dealing with the passage of light from a less dense to a more dense medium, n is greater than unity; and since the sine of any angle cannot be greater than unity, the quotient $\sin i/n$ cannot be greater than unity, and hence a value of r can always be found which will satisfy equation (a). That is, whatever the value of the angle of incidence (i) we obtain a refracted ray.

If, however, we are considering the passage of a ray of light from a more dense to a less dense medium, say from water to air, the refractive index, n, is less than unity. Since by equation (120), $n = 1/n'$ where n' is the refractive index from the less to the more dense medium, equation (a) may be written

$$\sin r = n' \sin i \quad . \quad . \quad . \quad . \quad (b)$$

where n' is greater than unity. The values of the angle of incidence i for which we obtain a refracted ray are now limited to values less than

that for which n' sin $i=1$, since for greater values of i the product n' sin i will be greater than unity; and as no angle has a sine greater than unity, no value of r can be found which will satisfy equation (b). When n' sin $i=1$, sin $r=1$, and hence $r=90°$, *i.e.* the refracted ray just grazes the surface of separation between the media. For larger angles of incidence there is no refracted ray, so that none of the light passes out of the denser medium, it all being *reflected* at the surface of separation according to the ordinary laws of reflection. The angle of incidence, of which the sine is equal to the reciprocal of the refractive index (reckoned from the less to the more dense medium), is called the *critical angle.*

If the critical angle (c) between two media is measured, we can obtain the refractive index from the relation

$$n=\frac{1}{\sin c} \qquad . \quad . \quad . \quad . \quad (122)$$

In the case of water and air the refractive index is 1·33, and hence the critical angle is given by

$$\sin c=\frac{1}{n}=\cdot 752.$$

$$\therefore \qquad c=48° \ 45'.$$

Thus a ray of light incident at a surface separating water and air at an angle greater than 48° 45′ will be entirely reflected, none of the light passing out into the air. A fish will see the whole of the landscape above the surface in a cone of which the semi-vertical angle is 48° 45′, and round this he will see a reflected image of the surrounding objects which are below the surface of the water.

112. Refraction through a Prism. Measurement of Refractive Index.—A portion of a transparent medium, such as glass, bounded by two plane faces which include a finite angle, *i.e.* are not parallel, is called a *prism.* A plane at right angles to both surfaces, and therefore also at right angles to the edge where the faces meet, is called a principal plane, and the angle between the lines where this plane cuts the faces is called the angle of the prism.

Since a principal plane is perpendicular to both faces, a ray of light incident on one face in a principal plane will continue in this plane throughout its passage through the prism.

When a ray of light passes through a plane slab the emergent ray is parallel to the incident ray. When we are dealing with a prism this is not the case, and the angle between the incident and emergent rays is called the angle of *deviation.*

If the angle of incidence of the incident ray with the first face is altered, it is found that for one angle of incidence the angle of deviation produced is a minimum, the deviation being greater both for smaller and larger angles of incidence. The angle through which the ray is deviated under these conditions is called the angle of minimum deviation.

Angle of minimum deviation.

If a ray of light PL is incident at such an angle on the face AB of the prism ABC, Fig. 192 (*a*), that the deviation is a minimum, the path of the ray in the prism is such that AM is equal to AL. If this were not so, let us suppose that PLMR, Fig. 192 (*b*), represents the path of a ray when the deviation is a minimum. Then a ray of light incident along RM would travel along RMLP, and hence would also suffer minimum deviation, for if we reverse the direction of a ray of light, it always retraces its path.

Next take the point l, such that Al = AL, and m, such that Am = AM,

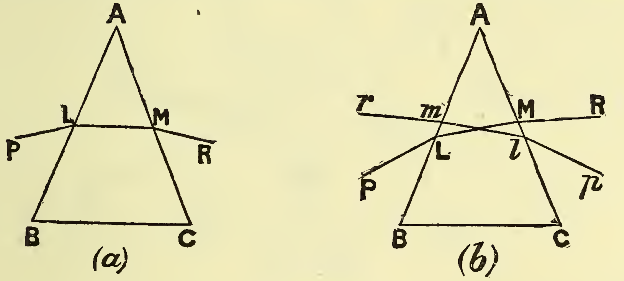


FIG. 192.

and join lm. Then draw pl inclined at the same angle to AC as is PL to AB, and mr inclined to AB at the same angle as is MR to AC. Then the path $plmr$ is exactly similar to the path PLMR, and hence a ray incident along pl would travel along $plmr$, and would be deviated through the same angle as is the ray PLMR, that is, it would suffer minimum deviation. Hence there are two rays, pl and RM, at different angles of incidence, both of which undergo *minimum* deviation, which is impossible, since by experiment there is only one angle of incidence which fulfils this condition. Hence PLMR cannot be at minimum deviation. In the same way, it can be shown that no ray which does not cut the two faces so as to make AL equal to AM can be at minimum deviation.

From the knowledge of the refracting angle of a prism and the angle of minimum deviation, we can calculate the refractive index from the medium surrounding the prism to the medium of which the prism is composed. Since in practice we have almost always to consider the passage of light from *air* into some other medium, we shall in future

refer to the refractive index from air into a medium simply as *the* refractive index of the medium.

Let ABC (Fig. 193) be the trace of a prism of which the refracting angle is A, the paper being a principal plane, and PLMR the path of a ray which is at minimum deviation, so that AL = AM. At L and M draw the normals NLN″ and N′MN″, also produce the direction of the emergent ray MR back to F, and produce the direction of the incident ray PL to cut this at E. Then the angle FEL or D is the angle through which the ray is deviated, and hence D is by supposition the angle of minimum deviation.

Since AL = AM the angle ALM is equal to the angle AML, and hence as the angles ALN″ and AMN″ are each a right angle, the angle N″LM is equal to the angle N″ML. In the quadrilateral ALN″M, the angles ALN″ and

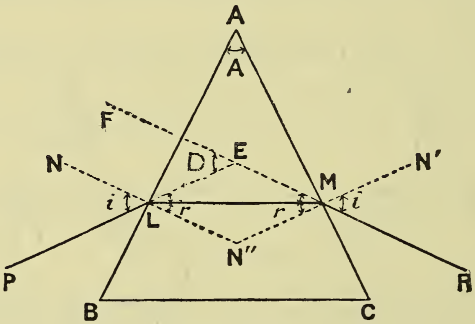


FIG. 193.

AMN″ are right angles, hence the angle LN″M = $\pi - $A. But since LMN″ is a triangle, the angle LN″M = $\pi - 2r$. Hence

$$\pi - 2r = \pi - A$$

$$\therefore \qquad r = \frac{A}{2}$$

In the triangle ELM the angles ELM and EML are each equal to $i - r$, and hence the exterior angle FEL is equal to $2(i-r)$, or D = $2(i-r)$,

$$\therefore \qquad i = \frac{D}{2} + r = \frac{D+A}{2}$$

But if n is the refractive index,

$$n = \frac{\sin\ i}{\sin\ r} = \frac{\sin\ \frac{1}{2}(D+A)}{\sin\ \frac{1}{2}A} \qquad . \quad . \quad . \ (123)$$

and so if the angles A and D are measured, the refractive index can at once be calculated.

The quantities A and D can both be measured by means of an instrument called a spectrometer. This instrument consists of a graduated circle ABC (Fig. 194 (*a*)), having a small astronomical telescope FG attached to an arm which can rotate round the centre of the circle, the position of the telescope being read by means of a vernier V. A tube DE, called the collimator, is fixed radially to the circle, and has a narrow vertical slit D at one end and a convex lens at the other. The distance between the slit and the lens is equal to the focal length of the lens, so that when the slit is illuminated by a source of light L, the rays of light that leave the lens E are all parallel.

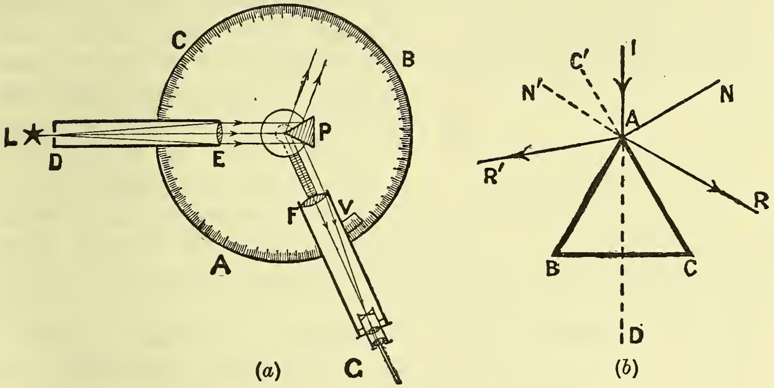


FIG. 194.

The prism P is placed on a small table attached to the circle, and when the refracting angle is being measured, the refracting edge is turned towards the collimator as shown in the figure. The telescope is then turned till the image of the slit, seen by reflection from one face of the prism, coincides with the intersection of two fine cross-wires placed in the eye-piece, and the vernier is read. The telescope is then turned till the image formed by reflection from the other face coincides with the intersection of the cross-wires, and the vernier again read. The difference between the vernier readings is equal to twice the angle of the prism.

Measurement of the angle of a prism and angle of minimum deviation.

Let IA (Fig. 194 (*b*)) be the direction of the light which falls on the prism when the angle is being measured, and AR and AR′ be the directions of the reflected rays. Then if the line CA is produced to C′ and AN is the normal to the face AC, we have that the angle of incidence IAN is equal to the angle of reflection NAR, and hence the angle IAC′ is equal to the angle RAC. But since IAD and C′AC are intersecting straight lines, the angle IAC′ is equal to the angle DAC so that this angle is equal to

the angle RAC. Hence the angle RAD is twice the angle DAC. Similarly the angle R′AD is twice the angle BAD. Thus the angle measured on the spectrometer, namely R′AR, is twice the angle BAC.

In order to determine D the prism is removed, and the vernier reading obtained when the telescope is turned so as to see the slit direct, thus obtaining the reading for the direction of the incident rays. The prism is then placed so that the light falls on one of the faces, and the telescope turned so as to catch the deviated rays. By turning round a vertical axis a position of the prism is found such that if it is turned in either direction, in order to catch the deviated light, the telescope has to be rotated towards the collimator, or, in other words, the prism is set at minimum deviation by trial. When this has been done, the vernier is read when the deviated image coincides with the cross-wires, and the difference between this reading and that for the direct light gives D.

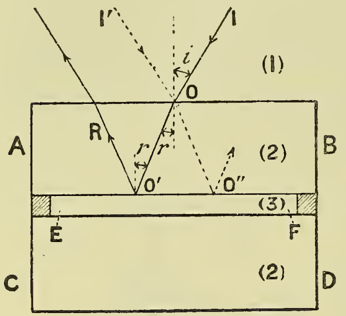


FIG. 195.

A very convenient method of measuring the refractive index of a substance, particularly for liquids, or when only a small quantity can be procured, depends on the measurement of the critical angle (§ 111) at which total reflection begins. Suppose we have two plates of glass AB and CD (Fig. 195) fixed together in such a way that they include a film of air EF, and that the whole is immersed in a fluid. Let us call the fluid medium 1, the glass medium 2, and the air medium 3. Then if a ray of light travelling in medium 1 is incident at an angle i on the glass, the reflected ray OO′ will be inclined at angle r to the normal. Since AB is a parallel-sided plate, the ray will be incident on the face separating the glass from the air at an angle r.

Measurement of refractive index by total reflection.

If now AB and CD are turned round, *i.e.* the angle of incidence i is altered till the ray OO′ is just totally reflected at O′, that is, till r is the critical angle from glass to air, we have the following relations:—

$$\frac{\sin i}{\sin r} = {}_1n_2,$$

$$\sin r = \frac{1}{{}_3n_2} = {}_2n_3.$$

Now $${}_1n_2 \times {}_2n_3 \times {}_3n_1 = 1.$$

Hence substituting $\dfrac{\sin i}{\sin r} \cdot \sin r \cdot {}_3n_1 = 1,$

or $\sin i \cdot {}_3n_1 = 1.$

Hence $${}_3n_1 = \frac{1}{\sin i}. \qquad (124)$$

But ${}_3n_1$ is the refractive index from medium 3 to medium 1, *i.e.* from air to the liquid, which is what we have called the refractive index of the liquid. Hence if we can measure i we can calculate n.

To determine the angle i, a parallel beam of light is allowed to traverse a glass-sided trough containing the liquid, and the glass plates, which are attached to a divided circle, are rotated till the light is no longer transmitted through the air film. The position of the plates is then read on the circle, and they are turned in the opposite direction, till the light is again totally reflected, *i.e.* till the light is incident along I′O. The difference between the new reading and the previous one gives twice the angle i.

113. Lenses. — A portion of a transparent refracting medium, bounded by two surfaces, one of which is spherical and the other is either plane or spherical, is called a *lens*.

If the two surfaces of the lens are spherical, the line joining the centres of the spheres is called the axis of the lens; if one of the surfaces is plane, the axis is the line drawn through the centre of the sphere perpendicular to the plane.

If the rays proceeding from a point P, after refraction at the lens, pass through a point P′; or if, although they do not actually pass through this point, their directions pass through P′, then P and P′ are called conjugate foci or conjugate points.

Each ray on one side of the lens which passes through P and suffers refraction through the lens, after it has traversed the lens, will have such a direction that it passes through P′. Thus to every ray on one side of the lens there is a corresponding ray on the other side of the lens, one ray being the continuation of the other. Such a pair of rays are called *conjugate rays*.

Lenses are divided into two classes. The first class, called convex lenses or converging lenses, are such that when a pencil of rays parallel to the axis passes through the lens, they are refracted so as to pass through the principal focus. The second class, called concave or diverging lenses, are such that when a pencil of rays parallel to the axis passes through the lens they are refracted, so that although they do not actually pass through the principal focus, yet their directions pass through the focus.

It is only when the medium of a lens is, as is generally the case, denser than the surrounding medium that the above definitions hold. In the opposite case a concave lens is a converging lens, and *vice versâ*.

In Fig. 196 are given the sections by a plane containing the axis of the three typical forms of convex lens. Lens (*a*) is called a double convex lens, (*b*) a plano-convex lens, and (*c*) a convexo-concave lens or convex meniscus. Lens (*c*) has one convex and one concave surface; the radius of the convex surface is, however, less than that of the concave. In Fig. 197 the three typical forms of concave lenses are shown. Lens (*a*) is called a double concave lens, (*b*) a plano-concave, and (*c*) a concavo-convex or concave meniscus. In (*c*) the concave surface has a smaller radius of curvature than the convex surface.

The point through which pass the rays, or their directions, which are conjugate to a series of rays parallel to the axis, is called a *principal focus* of the lens. In the case of a thin lens the distance between the principal focus and the lens is called the *focal length* of the lens. Every lens has two principal foci, one on either side of the lens, and if the media on the two sides of the lens are the same the two focal lengths of a lens are equal.

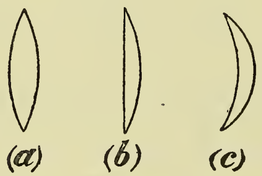


FIG. 196.

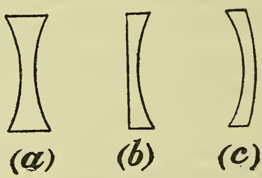


FIG. 197.

If a pencil of rays, all parallel to the axis, falls on a convex lens, such as AB (Fig. 198), after refraction through the lenses they all pass through the principal focus F, and OF is the focal length of the lens. Since the focal length, measured *from* the lens, is in the *same* direction as that in which the incident light is proceeding, it is negative, the same convention as to sign being adopted as in the case of mirrors (§ 110).

If a pencil of rays parallel to the axis falls on a concave lens, such as CD (Fig. 198), after their passage through the lens the rays diverge and travel as if they came from the principal focus F′, the point F′ being on the side of the lens on which the light is incident. Hence as OF′ is measured in the opposite direction to the incident light it is positive, so that the focal length of a concave lens is positive. It is for the above reasons that convex lenses are sometimes called negative lenses, while concave lenses are called positive; although the opposite convention in which a convex lens is called positive is generally adopted by opticians.

If r_1 is the radius of curvature of the first surface on which a parallel beam of light is incident, r_2 the radius of curvature of the other surface, f the focal length, and n the refractive index of the medium of which

the lens is composed, then these quantities, due regard being paid to their proper sign, are connected by the equation—

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad (125)$$

In the case of a double convex lens the centre of the sphere of which the first surface is a part is on the side of the surface remote from the source of light, hence the radius of curvature of this face is measured in the same direction as that in which the light is proceeding, and is therefore, according to our convention, to be reckoned as negative. The radius of curvature of the second surface is in the same way positive. In a double concave lens the radius of curvature of the first surface is positive and that of the second surface is negative.

Although the signs of the radii of curvature of the faces depend on

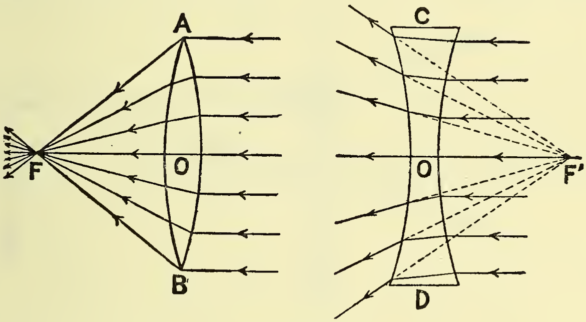


FIG. 198.

the *direction* of the light through the lens, the value obtained for the focal length by equation (125) above is independent of the direction. Thus suppose we have a double convex lens for which the radii of curvature of the faces are 10 cm. and 8 cm. If the light is incident on the 10 cm. face, so that the radius of curvature is negative, we get, if the refractive index of the glass is 1·5,

$$\frac{1}{f} = (1\cdot5 - 1)\left(-\frac{1}{10} - \frac{1}{8}\right) = \cdot5(-\cdot1 - \cdot125)$$

$$= -0\cdot1125$$

or $\qquad f = -8\cdot89$ cm.

If the light is incident on the 8 cm. face we have

$$\frac{1}{f} = (1\cdot5 - 1)\left(-\frac{1}{8} - \frac{1}{10}\right)$$

which obviously gives the same value for f as before.

The position and size of the image formed by a lens can be found by geometrical construction exactly similar to that used in regard to mirrors in § 110.

We take one ray which, proceeding from the point Q (Fig. 199) of

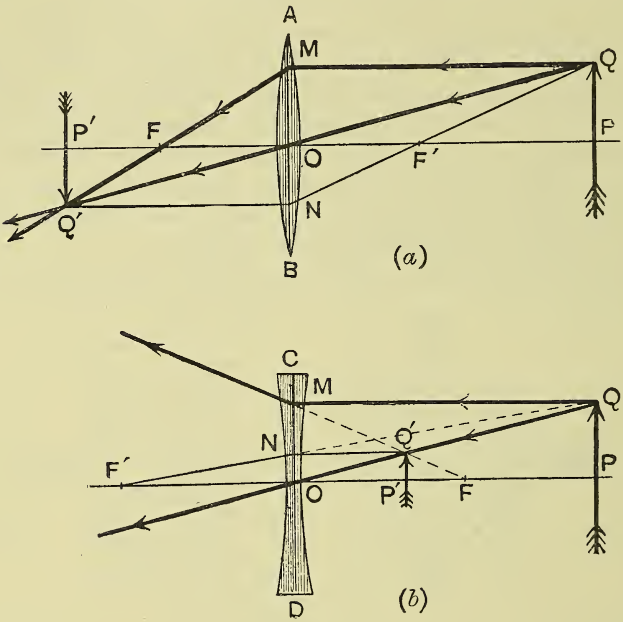


Fig. 199.

the object, passes through the centre of the lens O, and hence is undeviated, and another, QM, which is parallel to the axis, and hence after passing through the lens either actually passes, or its direction passes through the principal focus F. The point Q′ where the two rays intersect is the image of Q. If the rays actually pass through Q′ the image is real; if the rays do not actually pass through Q′, but only their directions when produced backwards, then the image is virtual. In Fig. 199, the image produced by the lens AB is real, while that produced by the lens CD is virtual.

Geometrical construction for finding position and size of image.

In the figure the triangles QOP and Q′OP′ are similar. Hence

$$QP/Q'P' = OP/OP' = u/v.$$

Thus the size of the object is to the size of the image as the distance of the object from the lens is to the distance of the image from the lens.

Since object and image are always interchangeable, that is, if P′Q′

were an object the image would be PQ, if we draw (Fig. 199) a ray Q′N parallel to the axis, this ray after refraction will pass through the principal focus F′ and meet the ray Q′OQ, which passes through the centre of the lens at Q.

In the case of both convex and concave lenses the triangles Q′P′F and MOF are similar, for the angles at F are equal and the angles at P′ and O are right angles. Hence

$$\frac{P'Q'}{OM}=\frac{P'F}{OF}$$

also the triangles QF′P and NF′O are similar, hence

$$\frac{ON}{PQ}=\frac{OF'}{PF'}$$

But since OM = PQ and ON = P′Q′

$$\frac{P'Q'}{OM}=\frac{ON}{PQ}$$

Hence
$$\frac{P'F}{OF}=\frac{OF'}{PF'}$$

But OF = OF′ = f. Hence if we call the distance PF′ U, and the distance P′F V, we get

$$\frac{V}{f}=\frac{f}{U}$$

or
$$UV=f^2 \quad . \quad . \quad . \quad . \quad . \quad (126)$$

Taking first the case of the convex lens AB (Fig. 199).

From the figure it is evident that PF′ = OP − OF′, and P′F = P′O − OF. But OP is equal to u and OF′ is equal in magnitude to f. Since, however, f for a convex lens is negative, although u is positive, we have

$$PF'=U=u+f.$$

Similarly, since v is negative

$$V=-v+f.$$

Substituting these values for U and V in equation (126), we get

$$(u+f)(f-v)=f^2$$

or
$$uf-vf=uv.$$

Dividing all through by uvf we get

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f} \quad . \quad . \quad . \quad . \quad (127)$$

Next taking the case of the concave lens CD. Here PF′ = PO + OF′ and

P′F = OF − OP′. Here, since u, v, and f are all positive and PO $= u$, OF $=$ OF′ $= f$ and OP′ $= v$, we have

$$PF' = U = u + f \text{ and } V = f - v.$$

Substituting these values in (126) we obtain

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

the same expression that we obtained in the case of a convex lens.

From the above equation, or by drawing a figure similar to that given for a mirror in Fig. 187, it can be shown that in the case of a convex lens if the object is situated between infinity and twice the focal length from the lens the image is real, inverted, and diminished, and is at a distance from the lens somewhere between f and $2f$, and on the opposite side to the object. If the object is distant between f and $2f$, the image is real, inverted, and magnified, and at a distance on the opposite side of the lens between $2f$ and infinity. If the object is between the principal focus and the lens, the image is virtual, erect, and magnified, and is situated on the same side of the lens as the object, being at a distance somewhere between 0 and ∞.

In the case of a concave lens the image is always virtual, erect, and diminished.

Power of a lens.

If the focal length, f, of a lens is expressed in *metres*, the quantity $1/f$ is called by opticians the *power* of the lens, the unit, that is, the power of a lens of which the focal length is a metre, being called a *dioptre*. The power of a convex lens is generally taken as positive. Thus a convex lens of which the focal length is 25 cm., that is, ·25 of a metre, would be 4 dioptres for $1/{\cdot}25 = 4$. Similarly a concave lens of which the focal length is 20 cm. has a power of -5 dioptres.

With the very long focus lenses used for spectacles it is sometimes difficult to tell by inspection whether a lens is convex or concave. If, however, the lens be held near the eye and is then moved to one side, if the image of surrounding objects seen through the lens appears to move in the *same* direction as the lens, the lens is *concave*. If the image appears to move in the *opposite* direction to that in which the lens is moved, the lens is *convex*.

When the lens is held near the eye any image seen clearly must be a virtual image (see § 114), and hence the object in the case of the convex lens must be at a distance from the lens less than the focal length. Let P be an object a little off the axis of the lens CD (Fig. 200), then the image will be at Q, the position being found by the construction given on

p. 323. If the lens be now moved down to C′D′, so that the axis is now A′B′, the image of P is at Q′, and an examination of Fig. 200 (*a*) will show that in the case of the convex lens Q′ is above Q, while Fig. 200 (*b*) shows that in the case of a concave lens Q′ is below Q, that is, for a convex lens the image appears to move in the opposite direction to the lens, and for a concave lens in the same direction.

The above method allows of a convex and concave lens of equal focal length being easily selected, for this combination when placed close together and tested as above gives no movement of the image. This is the principle of the method adopted by opticians for roughly determining the focal length of a lens. They have a selection of both convex and concave lenses of known focal lengths, and they pick out by this method a lens which, with the lens to be measured, gives no displacement. Then the focal length of the unknown lens is numerically equal to that of the one with which it is combined, but is of opposite sign.

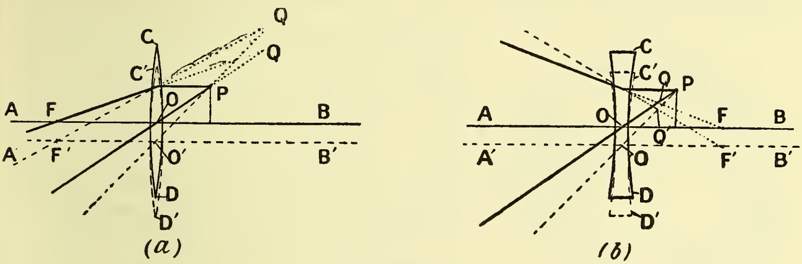


Fig. 200.

The power of a combination of two thin lenses placed in contact is equal to the algebraical sum of the powers of the individual lenses. Thus a convex lens of 25 cm. focal length (4 dioptres) placed in contact with a concave lens of 20 cm. focal length (-5 dioptres) forms a combination which is equivalent to a single concave lens of which the power is -1 dioptres, *i.e.* a concave lens of 100 cm. focal length.

Power of combination of two lenses placed in contact.

It is only when the aperture, that is the angle subtended by the lens at the principal focus, is small that all the rays from a luminous point situated on or near the axis which pass through the lens converge to a *single* point, *i.e.* form a point image. If the aperture is large, just as in the case of spherical mirrors, the rays which pass through the marginal portions of the lens do not intersect at the same point as that at which the rays passing through the central parts of the lens intersect, that is, we get spherical aberration. If a luminous point is some distance to the side of the axis of a lens the

refracted rays will not pass through a single point. They will pass through two short lines, called focal lines. These lines are separate and at right angles to one another. Between the focal lines all the rays pass through a small more or less circular area, called the circle of least confusion, and this forms the nearest approach to a point image of the luminous point, which can be obtained when the lens is inclined to the line joining the object and the centre of the lens.

CHAPTER III

OPTICAL INSTRUMENTS

114. The Eye.—The eye consists practically of a system of lenses by means of which a real image of external objects is formed on a network of nerves, called the retina, at the back of the eye, which nerves convey the impression of vision to the brain.

A diagrammatic section of the eye is shown in Fig. 201. The eye is surrounded, except in front, by a horny opaque coat, the sclerotic. The front transparent portion of this outside coating is called the cornea, C. The inside of the eye is divided into two portions by the iris I, the crystalline lens L, and the muscles which attach the latter to the walls of the eye. The crystalline lens is a double convex lens, of which the anterior surface has a radius of curvature of about 1·1 cm., while the posterior surface has a radius of curvature of about 0·8 cm. By means of the muscles attached to the edge of the lens the curvature of the faces, and hence the focal length, can be altered at will.

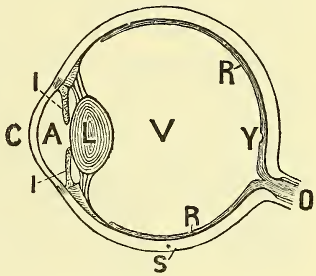


Fig. 201.

The iris I forms an opaque coloured diaphragm perforated by a central opening called the pupil. The diameter of the pupil varies with the intensity of the light which enters the eye; thus in a strong light the pupil is contracted, while in a feeble light it is expanded, these movements being involuntary. The space between the cornea and the lens is filled by a transparent liquid called the aqueous humour, while that between the lens and the retina R is filled with a liquid called the vitreous humour.

The retina consists of a semi-transparent network of nerve fibres formed by the spreading out of the termination of the optic nerve. Near the centre of the retina there is a round yellowish spot Y, called the yellow spot, and vision is most distinct when the image falls on this spot. The point of the retina where the optic nerve enters is insensitive to light, so that when the image of an object falls on this spot, which is called the blind spot, no sense of vision is produced.

If the eye really consisted of an ordinary lens, it is evident that for only one distance would the light from a luminous point be brought to a focus on the retina. At all other distances an indistinct and blurred image would be produced. The eye, however, possesses the power of accommodation so that the images of objects at very different distances can all be formed on the retina. In most cases the eye, when at rest, is so arranged that the image of a distant object is in focus on the retina. The accommodation for nearer objects is produced by a slight forward motion of the lens and an increase of the curvature of its surfaces, the increase in curvature of the front surface being very much the more strongly marked of the two.

The range of the accommodation is not unlimited, so that objects which are very near the eye cannot be clearly seen. Since when an object is a great way off we cannot make out small details about it, and neither can we do so when it is very near, it follows there must be some distance at which we are able to see most distinctly. The distance, which is called the *distance of distinct vision*, is for a normal eye between 25 and 30 cm., or 10 and 12 inches.

There are three defects of the eye which are of comparatively frequent occurrence. These are known as (1) short-sight, (2) long-sight, (3) astigmatism.

Defects of the eye.

In the case of short-sight, or myopia, distant objects cannot be seen distinctly, because the point to which the rays from distant objects are brought to a focus is, even when the lens is at its flattest, in front of the retina. We may here consider that the lens is too convergent for the size of the eye-ball, so that if in front of the eye we place a concave lens so as to make, with the lens of the eye, a less convergent system than the crystalline lens alone, the defect of short-sightedness can be corrected.

Since the image formed by a concave lens is always virtual, it is evident that if d is the *maximum* distance at which a short-sighted person can see distinctly, then if the concave lens is such that the focus for parallel rays is at a distance d from the eye, the eye will be able to see clearly this image, and hence all distant objects. Since the spectacle lens is always placed quite close to the eye, the distance of the focus for parallel rays from the lens must be d, that is, the lens must have a focal length d.

In long-sight, or hypermetropia, near objects cannot be seen distinctly, this being due to the fact that, when the lens is as much curved as possible, the image of objects even some distance off is formed behind the retina. In this case the eye, when relaxed, is in such a state that parallel rays meet behind the retina, so that to see distant objects the eye has to be accommodated. This defect can be remedied by placing

a convex lens in front of the eye, for by this means the focus of the combination of lens and eye is nearer the crystalline lens than when no spectacle lens is used.

Let the minimum distance at which a long-sighted eye can see clearly be d, and it be required to find the focal length of a convex lens which will produce distinct vision at the ordinary distance of most distinct vision, say D. Then, assuming that the spectacle lens and eye are close together, we must take the focal length f of the lens such that the image produced by an object at a distance D must be on the *same* side of the lens as the object, and at a distance from the lens d, where $d > D$. Here $u = +D$ and $v = +d$. Hence

$$\frac{1}{f} = \frac{1}{d} - \frac{1}{D} \quad . \quad . \quad . \quad . \quad (128)$$

Since d is $> D$, $\frac{1}{d}$ will be less than $\frac{1}{D}$, and hence $\frac{1}{f}$, and therefore f will be negative, and the lens must be convex, which agrees with the conclusion at which we have already arrived.

In astigmatism the surfaces of the cornea, and the lens, but principally the former, are not symmetrical about the axis. In most cases the vertical section of the cornea of an astigmatic eye is more curved than a horizontal section, so that the image of a horizontal line is formed nearer the crystalline lens than the image of a vertical line. This defect is remedied by the use of spectacles in which the surfaces of the lenses are not spheres, but differ from these in the opposite sense to that of the defective eye.

Suppose AB and CD (Fig. 202) are two lenses, the distance between the lenses being equal to the focal length of the lens CD, which is convex. Then if PM be an object and from P we draw a ray PFE, through the centre of the lens AB, this ray will be undeviated by AB, and as it passes through F the principal focus of CD, after refraction its course EG will be parallel to the axis. Now this ray must of necessity pass through the image Q of the top of the object. Hence whatever the position of the image, *i.e.* the nature of the lens AB, the height of the image is constant, that is the magnification is constant, and is the same as if the lens AB were absent. If CD represents the lens of the eye and AB is a spectacle lens, we therefore see that if the spectacle lens is placed at a distance in front

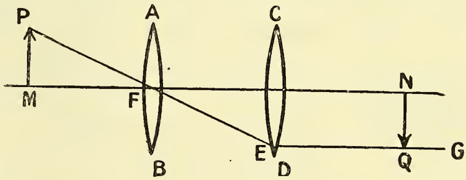


FIG. 202.

Effect of spectacles on the size of image.

of the eye equal to the focal length of the eye, the size of the image formed on the retina will not be affected by the spectacles. If the spectacles are further from the eye than the principal focus (it is not generally practicable to get them nearer), with convex glasses the size of the image is increased. With concave glasses the contrary is the case. Hence persons who use concave spectacles generally wear them close to their eyes, while those who wear convex spectacles sometimes put them near the end of their noses, so as to obtain a large image.

115. The Microscope.—We have seen in the preceding section that if we attempt to increase the distinctness with which an object can be seen by bringing it nearer the eye, so that it appears larger, a position is at length reached such that if we bring it nearer we are unable to see it at all distinctly.

If a convex lens is placed in such a position that the object AB

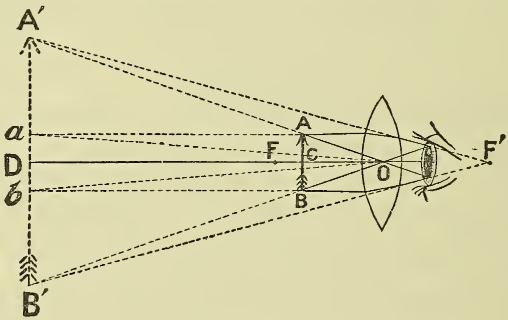


FIG. 203.

(Fig. 203) is between the principal focus F and the lens, the rays, after they leave the lens, will proceed as if they came from the virtual image A′B′. This image is found by the construction given in § 113, and is erect and magnified. Now, although the image A′B′ and the object AB subtend nearly the same angle at the eye, yet if A′B′ is at the distance of distinct vision, on removing the lens the object AB would be seen very indistinctly, it being so much within the minimum distance of distinct vision. Hence, to find the magnification produced by the lens, we must compare the angle subtended at the eye by the image A′B′ with the angle subtended by the object when it is at the distance at which it can be most clearly seen, *i.e.* at *ab*. If the eye is very near the convex lens, the angles subtended by the image, and by the object *when at the distance of distinct vision*, are very nearly equal to A′OB′ and *aob* respectively.

The simple microscope or magnifying glass.

Hence the magnification is equal to $\frac{A'B'}{ab}$ or $\frac{A'B'}{AB}$, or, since the triangles

A′OD, AOC are similar, to $\frac{DO}{CO}$. But CO is the distance of the object from the lens, and DO is the distance of the image from the lens, hence we have the magnification $= \frac{v}{u}$. If f is the focal length of the lens,

$$\frac{1}{f} = \frac{1}{v} - \frac{1}{u},$$

or

$$\frac{v}{u} = 1 - \frac{v}{f} \quad . \quad . \quad . \quad . \quad (129)$$

Hence the magnification $= 1 - \frac{v}{f}$; or, since the image is to be formed at the distance of distinct vision D, the magnification $= 1 - \frac{D}{f}$, where it must be remembered that f is negative and D is positive.

From the above expression it will be seen that the magnification

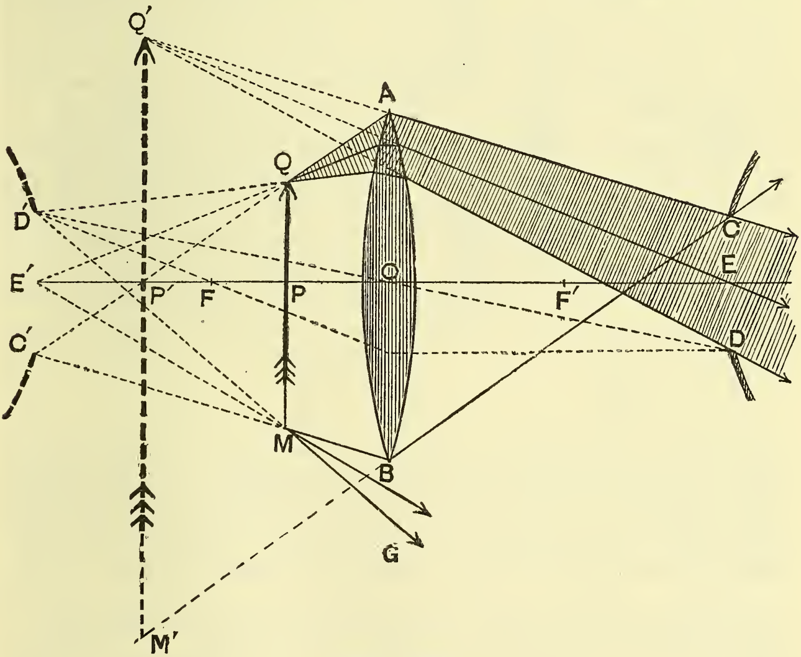


FIG. 204.

increases as f decreases, so that to obtain great magnifying power a lens of very short focal length must be taken. Since if a very short focus lens is employed, the aperture, *i.e.* ratio of diameter to focal length, is large even for a small lens, spherical aberration is considerable. To reduce this effect in Wollaston's lens and the Coddington lens an opaque

stop is placed within the lens so that all the pencils of rays which reach the eye must pass through the central part of the lens.

The whole of the rays of light which leave any point of the object and strike the lens do not in general enter the eye, for the pupil is usually smaller than the cross-section of the pencil of rays, and so the iris acts as a stop. In Fig. 203, the limiting rays, proceeding from the points A and B of the object, which are able to enter the pupil are shown; all rays outside the cone included by these rays are stopped by the iris, or peripheral parts of the eye.

If CD (Fig. 204) represents the pupil of the eye, and if we find the image C′D′ of the pupil formed by the lens AB, then all the rays from any point of the object which actually enter the eye must proceed before they meet the lens, so that their directions pass through C′D′. This is at once evident, for a ray which proceeds from Q so that when produced backwards it would pass above D′ will, since D is the image of D′, after passage through the lens pass through a point below D, and hence be intercepted by the iris. In the figure the whole pencil of rays from the point Q which enters the eye is shown shaded. It will be noticed that Q has been taken so that the pencil just reaches the edge of the lens. In the case of points further from the axis than Q, only part of the pencil will strike the lens, so that on this account fewer rays will reach the eye than is the case for points between Q and P. At M, which lies on the line joining C′ to the edge B of the lens, only the extreme ray reaches the eye, while for points outside M, no ray reaches the eye, and hence such points are outside the field of view. If, however, the diameter of the lens were increased, say to G, then rays from points further from the axis than M would be able to enter the eye, and so the field of view would be extended.

Since the edge of the lens limits the extent of the field of view, it is called in this case the field of view stop. The image C′D′ of the pupil of the eye formed by the lens is called the entrance-pupil, for all rays which after refraction at the lens enter the pupil of the eye, called the exit-pupil, must, before meeting the lens, either pass through the entrance-pupil or their directions must pass through the entrance-pupil.

Owing to spherical aberration the greatest magnification that can be obtained with a single lens is about one hundred-fold. When greater magnifications are required, recourse is had to a combination of lenses forming what is called a compound microscope. In its simplest form

The compound microscope.

the compound microscope consists of two convex lenses, A and B (Fig. 205). The lens A, or objective, is of short focal length, and is so placed that the object PQ is just beyond its principal focus, so that a real, inverted and slightly magnified image is produced at P′Q′. The second lens or eye-piece, B, is placed at

such a distance from the objective that the image formed by the latter is just inside the principal focus F, and hence the eye-piece, acting as a simple microscope, gives a virtual and magnified image $P''Q''$.

Since the normal eye, when at rest, is adjusted for an object at a great distance, *i.e.* for parallel rays, it is less trying to the eyes if, whenever possible, we arrange an optical instrument so that the rays that enter the eye are parallel. In the case of the compound microscope this can be easily done, for if the lens B be placed so that the image $P'Q'$ is formed at its principal focus, then the rays from each point of the image $P'Q'$, after passing through the eye-piece, will emerge parallel. The angle

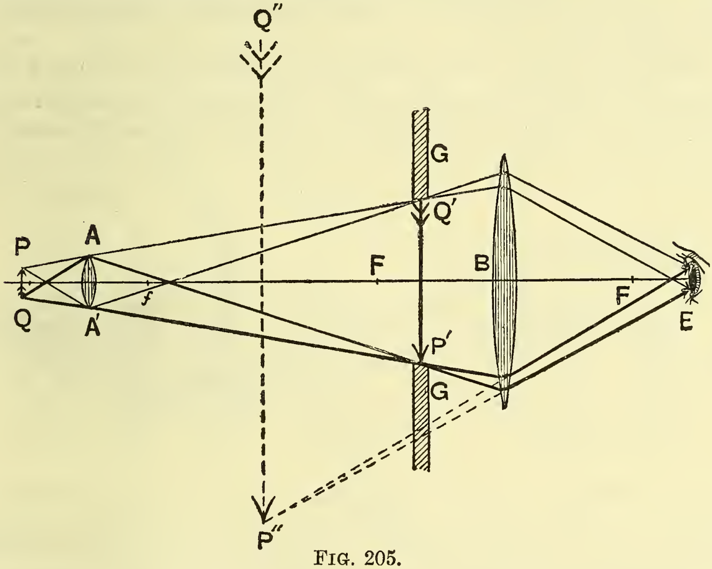


Fig. 205.

between the rays from the extreme points of the object when they enter the eye remains practically the same, as in Fig. 205, so that the apparent size of the image remains the same; we have, however, done the focussing by means of the instrument, instead of using the accommodation of the eye.

The objective and eye-piece lenses are generally chosen of such sizes that the whole of the rays from any given point of the object which strike the objective are transmitted to the eye. Thus the edge of the objective acts as the limit to the entrance-pupil. In the absence of any diaphragm the edge of the eye-piece lens would limit the field of view. A diaphragm is, however, generally introduced at G, the aperture being of such a size that the full pencil of rays from a point of the object, such as P, the image of which coincides with the edge of the diaphragm, will meet the lens B.

In this case the edge of the field of view is sharply defined, and the illumination is the same all over the field. If the edge of the lens B acts as the field of view stop, the illumination at the edge of the field gradually falls, owing to the pencils corresponding to the outside portions of the field partly failing to meet the lens B. This result is at once evident if we compare Fig. 204. The exit-pupil is at E, which, as is evident from the figure, is the image of the lens AA′ formed by the lens B. When using the instrument the eye is placed so as to be at the exit-pupil. The diameter of the exit-pupil is generally less than the diameter of the pupil of the eye. If it is not, then the pupil of the eye acts as exit-pupil, and the image of the pupil of the eye formed by the instrument acts as the entrance-pupil.

The above form of microscope is much simpler than any now used, but to go into the theory of the modern microscope to any purpose would be beyond the scope of this work. On account of spherical aberration

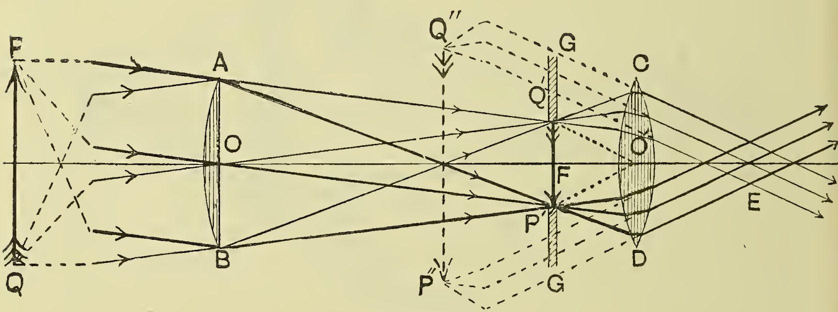


FIG. 206.

and chromatic aberration, both the objective and the eye-piece consists of combinations of lenses, the objectives of some modern high-power microscopes having as many as ten separate lenses.

116. Telescopes.—A telescope is an instrument for obtaining a magnified image of a distant object. In its simplest form it consists of a convex lens of large focal length, called the object-glass, AB (Fig. 206). This lens forms a real, inverted image, P′Q′, of the object[1] at the principal focus. This image is just inside the principal focus of a short-focus convex lens CD, which produces an enlarged virtual image P″Q″ of P′Q′ in exactly the same manner as occurs in the simple microscope. This virtual image is the same way up as P′Q′, hence the image seen by the

[1] In the figure the object PQ is shown close to AB, but the dotted lines are supposed to be extended so much that the three rays shown leaving either of the points P or Q are to all intents and purposes parallel. Similarly the image P″Q″ is really at a great distance, but in order to indicate the fact that it is inverted it is shown on the figure.

eye is inverted. For astronomical purposes this inversion is of no consequence, and hence such an inverting telescope is generally called an astronomical telescope.

Magnifying power of telescope.

Since OQ′ and OP′ are straight lines drawn from the extremities of the distant object, the angle which this object would subtend at the eye is equal to P′OQ′, or in circular measure $P'Q'/F$, where F is the focal length of the object-glass. If the eye-lens is adjusted so that parallel pencils enter the eye, P′Q′ must be at the principal focus of the eye-lens; and it is evident from the figure that the angle subtended by the image seen is Q′O′P′, and the circular measure of this angle is $P'Q'/f$, where f is the focal length of the eye-lens. Hence the ratio of the angle subtended by the object seen with the naked eye to that subtended by the image, which ratio is called the magnification, is F/f. Thus the magnification of a telescope depends on the *ratio* of the focal length of the objective to that of the eye-piece. The diameter of the objective has no effect on the magnification, although owing to diffraction a shorter focus eye-piece can with

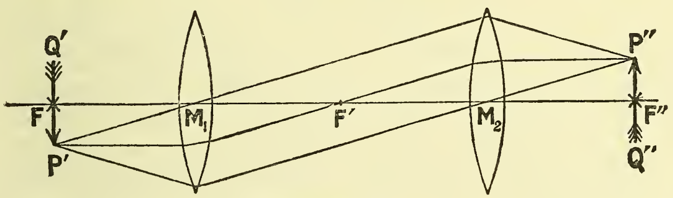


Fig. 207.

advantage be employed with a telescope having an objective of large diameter, and hence indirectly the diameter of the objective influences the greatest magnification which can be employed.

Field of view.

The edge of the object-glass generally acts as the entrance-pupil, so that all the rays proceeding from any given point of the object which strike the object-glass pass through the telescope. The image at E of the object-glass produced by the eye-piece is the exit-pupil, and the observer's eye is placed at E. If the pupil of the eye is smaller than the exit-pupil then some of the rays will be cut off by the iris, and the peripheral parts of the object-glass will not be operative, as can easily be seen by drawing a figure. A diaphragm G placed at the principal focus of the eye-piece acts as a field of view stop, its functions being the same as in the case of the microscope.

The image seen in an astronomical telescope is inverted, and although this does not matter for astronomical purposes, yet it would be very inconvenient when the telescope is used to view terrestrial objects.

An erect image is obtained by placing two convex lenses between the

object-glass and eye-piece. If these two lenses are at a distance apart equal to the sum of their focal lengths, the one nearer the objective is placed so that the image formed by the objective is at the principal focus of this lens. In Fig. 207 $P'Q'$ is the image formed by the objective, and as this image is at the principal focus of the lens M_1, all the rays leaving any point of the object will, after their passage through the lens M_1, be parallel. These parallel rays, falling on the lens M_2, form an image $P''Q''$ at the principal

Terrestrial telescopes.

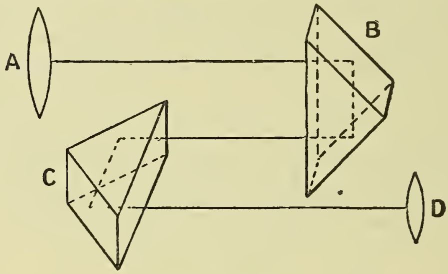


Fig. 208.

focus, F'', of this lens, which image is inverted with reference to $P'Q'$, and hence erect with reference to the object. This image, $P''Q''$, is viewed by an eye-piece as in the astronomical telescope. The two lenses M_1 and M_2 serve simply to give an erect image, and are fixed at a constant distance apart. The telescope is focussed by altering their distance from the object-glass, so that the image formed by the latter is always at the principal focus of M_1.

The objection to this method of obtaining an erect image is the

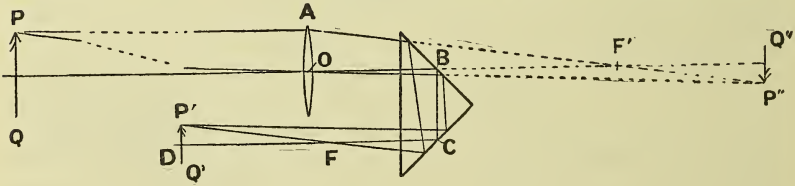


Fig. 209.

increase which it causes to the length of the telescope. Another way of obtaining an erect image and at the same time much reducing the distance between the object-glass and eye-piece is to reflect the rays twice during their passage between the object-glass and eye-piece in two prisms. The rays of light which pass through the objective A (Fig. 208) fall normally on the hypotenuse face of a right-angled prism B. When the rays reach the back face of the prism the angle of incidence is greater than the critical angle (§ 111), and hence the rays are totally reflected on

to the second back face, where again they are totally reflected, so that they finally emerge again through the hypotenuse face. The rays then fall on a similar right-angled prism C, which is placed with the edge which includes the right angle at right angles to the similar edge of B. Finally the rays enter the eye-piece D. The distance along the path ABCD is equal to the distance between the object-glass and eye-piece in the astronomical telescope.

If we consider the case of a single prism (Fig. 209), the optical axis

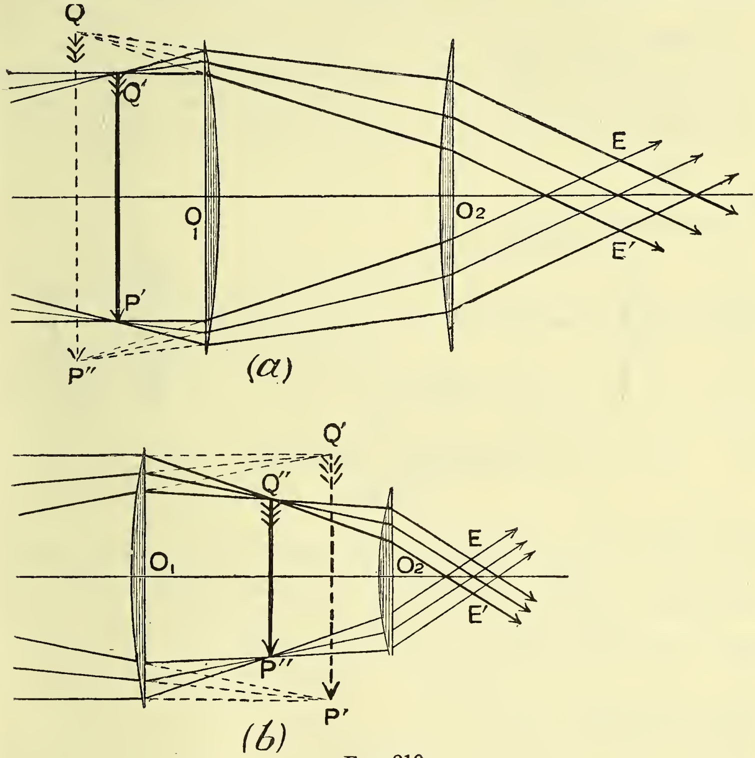


FIG. 210.

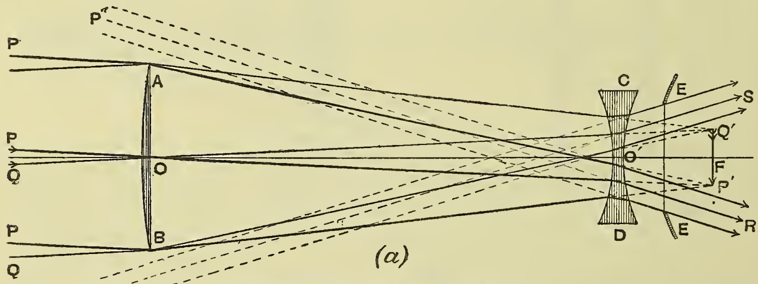
of the system being OBCFD. Then an object at PQ will give an image P′Q′ in the manner shown in the figure, and it will be observed that this image is the same way up as the object. It, however, differs from the object in that it is reversed, right for left. The reflection in a second prism placed at right angles to the first reverses the image right and left, and hence the final image is truly "erect."

If a single lens is used as the eye-lens of a microscope or telescope, a comparatively great focal length must be employed, otherwise the image is much disturbed, due to spherical aberration. The effects of

spherical aberration are much reduced if a compound eye-piece consisting of two lenses placed at some distance from one another is employed. There are two chief types of such compound eye-pieces.

Compound eye-piece.

In the *Ramsden or positive eye-piece* there are two plano-convex lenses of equal focal length placed at a distance apart which varies between the focal length of either and 2/3 of this distance. As shown at (*a*) Fig. 210, the curved sides of the lenses are turned towards each other. The object-glass of the telescope or microscope, as the case may be, forms the real image at P′Q′, the rays then strike the first lens O of the eye-piece, called the field lens, and proceed as if they came from the image P″Q″. After passing through the lens O_2, called the eye-lens, the rays either emerge parallel or as if they came from a virtual image at the distance of distinct vision.



MAGNIFICATION 5 DIAMETERS

FIG. 211.

In the figure the complete pencil of rays forming the two extreme points of the image are shown. The field of view stop is placed in the plane of P′Q′, and since this image is real cross-wires or a divided scale can be placed in this plane when it is desired to make micrometric measurements. The exit-pupil of the system is at EE′ and the pupil of the eye ought to be placed at this point.

The *Huyghens or negative eye-piece* is shown at (*b*) Fig. 210. It consists of two plano-convex lenses with their curved faces turned towards the objective. The focal length of the lens O_1 is three times that of the lens O_2, and the distance between the lenses is equal to twice the focal length of the lens O_2. In the absence of the eye-piece the object-glass would form a real image at P′Q′. On account, however, of the rays meeting the field lens O_1, a real image is formed at P″Q″. The rays then traverse the eye-lens O_2 and emerge in parallel pencils as shown. The field of view stop is placed in the plane of the image P″Q″, and the exit-pupil is at EE′. In the case of this eye-piece, since the object-glass does not form a real image, it is not possible to place

a cross-wire or micrometer scale in the plane of this image. If the cross-wires or micrometer is placed in the plane of P″Q″, the image formed by the objective may be distorted by the lens O_1, but the micrometer scale, being on the eye side of this lens, would not suffer such distortion. In the Ramsden eye-piece, however, any distortion will affect the image formed by the objective and the micrometer scale equally, for they are both observed through *both* lenses of the eye-piece.

A further advantage of the compound eye-piece is that it is practically achromatic (see § 118).

A form of telescope which gives an erect image with only two lenses, and which can be made much shorter than the terrestrial telescope described above, is Galileo's. This form of telescope consists of a convex lens as object-glass, AB (Fig. 211), and a concave lens, CD, as eye-piece. If CD were not present the convex lens would form a real image at P′Q′; when the concave lens is interposed between AB and the image, so that the distance O′F is equal to the focal length of CD, the rays of light from any one point of the object will be parallel after they leave this lens. Hence, as shown in the figure in thick lines, for a pencil of rays coming from the point P of the object, the rays will enter the eye in the direction P″R, while a pencil coming from Q will, as shown by the light lines, enter the eye in the direction Q″S, so that the eye sees an enlarged and erect image. The magnification is, as before, equal to F/f, where F is the focal length of the objective and f that of the eye-piece. In Galileo's telescope the distance between the objective and eye-piece is $F-f$, while in the astronomical telescope it is $F+f$, hence the saving in length for an equal magnifying power, and with objectives of equal focal length. Opera and field-glasses, other than prismatic, consist of two Galilean telescopes, one for each eye, the distance between the objectives and eye-pieces being variable by means of a screw, so that the image formed by the objective may always be formed at the principal focus of the eye-piece.

Galilean telescope.

In the photographic camera a system of lenses, which is equivalent to a convex lens, forms a real image in the plane of the photographic plate. The object of using several lenses in place of a single lens is firstly to reduce spherical aberration, so that a lens of wide aperture may be employed, and secondly to reduce chromatic aberration.

CHAPTER IV

DISPERSION

117. Dispersion.—The phenomenon of refraction is not in reality as simple as we have hitherto considered it to be, for if a narrow parallel pencil of white light, such as sunlight, is allowed to pass obliquely from one medium to another, it is found that in the second medium the white light is split up into light of several colours, a phenomenon which is referred to as *dispersion.*

Thus if a beam of parallel rays of white light, such as is obtained by reflecting sunlight through a narrow slit, is introduced into a dark room and meets a screen DE at F, forming a white patch of light, then on interposing a prism ABC (Fig. 212) in the path of the beam with its refracting edge parallel to the slit, the light will be refracted towards the base of the prism, but the patch on the screen is no longer the same size as before, nor is it white. The patch is drawn out in the direction RV, in which the light is deviated, and exhibits all the colours of the rainbow. These colours pass imperceptibly the one into the next, but starting with red nearest the original undeviated patch F, the colours pass through orange, yellow, green, blue, indigo, to violet, which is the most deviated. These colours constitute what is called a spectrum.

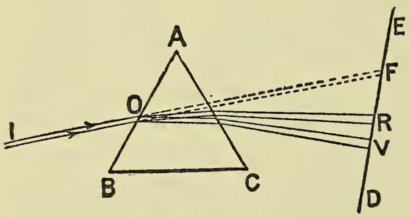


FIG. 212.

Thus white light has been split up by the prism into light of a number of different colours, these coloured lights being deviated to a different amount by the prism, so that the refractive index between two media, on which the deviation depends, is different for light of different colours; and since the violet rays are more deviated than the red, the refractive index for violet light is greater than for red light.

That white light is really formed by the superposition of light of all the colours of the spectrum can be shown by receiving the colours of the spectrum on a number of separate mirrors, and reflecting the light from

them to the same point, when it will be found that white light will be reproduced.

In the form of the experiment described above, the different colours overlap on the screen to a certain extent; and in order to obtain a spectrum where no overlapping takes place, or a *pure spectrum*, as it is called, we may adopt the arrangement shown in Fig. 213. Light from a source L, such as the electric arc, passes through a narrow slit in a screen S, and then falls on a convex lens A, which, when the prism is not interposed, forms a real image of the slit at S′. If now the prism is interposed at B, the light will be deviated towards the base of the prism, and a spectrum will be formed on a screen placed at D. If we suppose that the slit is illuminated by violet light only, then an image of the slit will be produced at V, while if red light is used the image will be at R. Hence the spectrum VR is composed of a series of images of the slit formed by differently coloured light. If the slit is very narrow, one image will overlap very little on the adjacent images, and a pure spectrum will be obtained. As the slit is widened the images will overlap more and more, till with a very wide slit we shall get a white patch in the centre of the spectrum where all the images overlap, with a red edge at one end and a violet edge at the other.

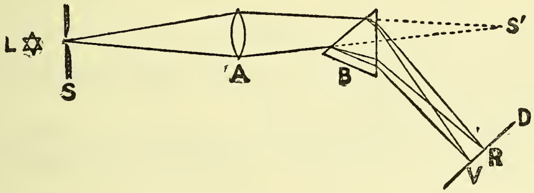


FIG. 213.

Methods of producing a pure spectrum.

Another method of obtaining a pure spectrum is shown in Fig. 214. Parallel light, which may be obtained by means of a collimator, being incident on the prism, a lens L is placed after the prism, and this lens brings the rays of the different colours to real foci between R and V, where a pure spectrum may be received on a screen or viewed with an eye-piece.

This arrangement will also allow of the recomposition of the different colours of the spectrum to form white, for if the screen be placed at AB the red and violet rays, as shown by the figure, and therefore also the rays of the other colours, will be uniformly spread over the patch AB. Under these conditions a white patch will appear on the screen. If, however, a small obstacle be placed at V, so as to cut off the violet, the patch at AB will appear coloured a greenish-gold colour, produced by the

mixture of the remaining colours. In the same way, by cutting off the red rays by an obstacle placed at R, the patch will appear a greenish-blue.

When the slit of the spectrometer shown in Fig. 193 is illuminated with white light, a pure spectrum is formed at the principal focus of the lens F in the manner considered above, and can be observed with the eye-piece. The spectrometer, when used to observe spectra, is sometimes called a spectroscope. By using light of different colours, the

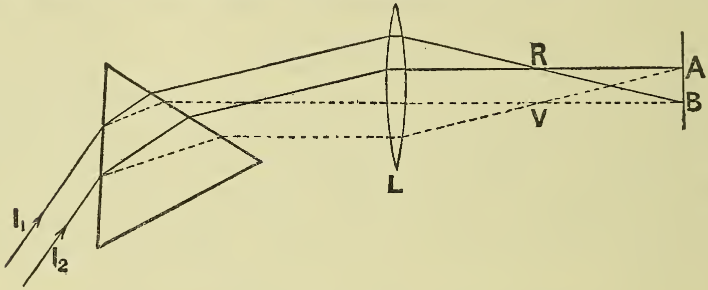


Fig. 214.

refractive index of a substance for light of these colours can be obtained by any of the methods given in § 112.

When the slit of a spectroscope is illuminated by sunlight, it is found that the spectrum is traversed by an enormous number of dark lines parallel to the length of the slit. These dark lines are called *Fraunhofer's lines*, and are due, as we shall see later, to the light of the colours which are thus missing from the solar spectrum being absorbed in the sun's or the earth's atmosphere.

These lines form a very convenient means of specifying any particular colour in the spectrum, and hence the more prominent of them

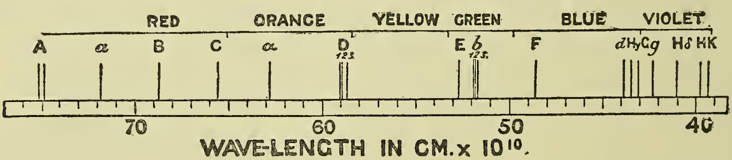


Fig. 215.

are indicated by the letters A, B, C, D, &c. Their relative position in the spectrum are shown in Fig. 215. The lines A, B, and C are in the red, D in the orange-yellow, E in the green, F in the greenish-blue, G in the indigo, and H in the violet part of the spectrum. Hence when we refer to light of any particular colour as, say, D light, we mean light of the colour which corresponds to the dark line D in the orange-yellow of the solar spectrum.

The angle included between the emergent rays of any two colours produced by a prism, that is the angle RAV (Fig. 214), is called the *dispersion* of the prism for these two colours. The ratio of the dispersion for two colours to the deviation for light of a colour half-way between the two is called the *dispersive power* of the material of which the prism is produced. If a prism is at minimum deviation we have (equation 123)

Dispersive power.

$$n = \frac{\sin \frac{1}{2}(D+A)}{\sin \frac{1}{2}A}$$

Hence if the angle of the prism is small, so that both A and D are small, this expression reduces to

$$n = \frac{A+D}{A}$$

or

$$D = A(n-1) \quad . \quad . \quad . \quad . \quad (130)$$

Hence if n_r is the refractive index of a given *thin* prism for red light, D_r the deviation for this light, and n_b and D_b the corresponding quantities for blue light, the dispersion, which is equal to the difference of the devitions, is equal to $D_b - D_r$ or $(n_b - n_r)A$. If n_d and D_d are the refractive index and deviation for a yellow, which is half-way between the red and blue, the dispersive power D of the material of the prism is given by

$$P = \frac{D_b - D_r}{D_d} = \frac{n_b - n_r}{n_d - 1} \quad . \quad . \quad . \quad (131)$$

The dispersive powers of materials differ greatly, as will be seen by the following table:—

Substance.	Dispersive Power.
Water	0·042
Carbon bisulphide	0·145
Crown glass	0·043
Flint glass	0·061
Rock salt	0·057

118. Achromatism.—Since the dispersive powers of different media are different, it follows that if we have, say, a crown-glass prism and a flint-glass prism, and adjust the angles of the prisms so that the dispersion between blue and red $(D_b - D_r)$ is the same, then the deviation for the mean colour D_d will be different. Thus if two such prisms are placed with their refracting angles in opposite directions, and a parallel beam of light is passed through the combination, the dispersion due to one will be neutralised by that due to the other. On the other hand, the deviation produced by the crown prism will be greater than that in the opposite direction produced by the flint prism. Hence the

combined prism gives deviation without dispersion, and is said to be *achromatic.*

If, on the other hand, the angles of the two prisms are chosen so that the deviation, D_d, of the mean colour is the same for both, the dispersion produced by the flint-glass prism will be greater than that produced by the crown-glass prism. Hence when the prisms are opposed, the combination will allow yellow light to pass through without deviation. The red will be slightly deviated towards the base of the crown-glass prism, and the blue slightly deviated towards the base of the flint-glass prism. Thus a spectrum will be produced, although the yellow is undeviated. This arrangment forms what is called a *direct vision spectroscope.*

In considering the formation of images by lenses, we have supposed that the light was monochromatic. When white light is used, we shall not only get the deviation which we have hitherto considered, but also dispersion.

Suppose we have a convex lens AB (Fig. 216), and that a parallel

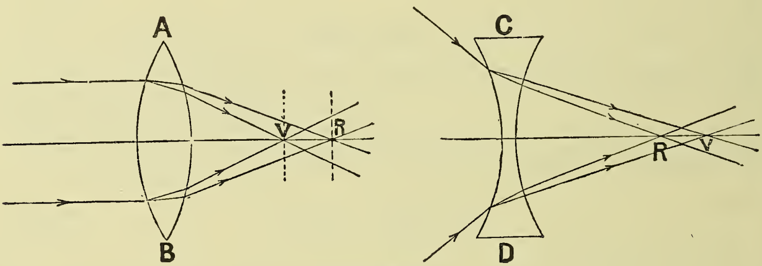


FIG. 216.

beam of white light falls on it, then where the rays enter and leave the lens the violet rays will be more deviated towards the axis of the lens than are the red rays, and hence the violet rays will be brought to a focus at a point V nearer the lens than the point R, where the red rays are brought to a focus. Thus if a screen is placed at V we shall get a central violet spot surrounded by a red ring, while if the screen is placed at R there will be a central red spot surrounded by a violet ring.

If a convergent pencil of rays is incident on a concave lens, CD, the violet rays are brought to a focus at a point V, which is *further* from the lens than the point R, where the red rays come to a focus.

If the convex and concave lens are of the same material and of equal focal length, on being placed close together the two dispersions counteract each other, but in this case the deviation would also be nil, and the whole would simply act like a plane slab. By making the convex lens of crown glass, and the concave lens of flint glass, we are, however, able, as in the case of the prisms, to obtain equal and opposite dispersion, and

still have deviation in the direction of that produced by the crown glass, *i.e.* the combination will be a convex lens, so that it is possible to construct achromatic lenses.

By the use of two lenses it is possible to make a lens which shall be achromatic as far as light of any two colours is concerned. The combination will not, however, be achromatic for light of other colours. If in place of two lenses we use three, made of materials having different dispersive powers, the combination can be made achromatic for light of three colours, and so on. The colours for which the lens system is rendered achromatic depend on the purpose for which the lens is to be used. Thus for a telescope used in eye observations the colours chosen are those parts of the spectrum which affect the eye most strongly, while if the telescope is to be used for photography, it is most important that the lens should be achromatised for the violet and ultra-violet, since these rays are chiefly concerned with the production of the photographic image.

It can easily be proved that the condition for achromatism is that the ratio of the focal lengths of the convex and concave lenses shall be equal to the ratio of the dispersive powers of the two glasses. (See Watson's *Text-Book of Physics*, § 374.)

CHAPTER V

LIGHT CONSIDERED AS A WAVE-MOTION

119. Velocity of Light.—In the preceding chapters we have dealt with some of the phenomena of light without considering what is the nature of the disturbance which when it enters the eye produces the sensation of light. We now proceed to consider certain phenomena which depend on the fact that light is propagated by a form of wave-motion. A consideration of the *nature* of this wave-motion is beyond the scope of this book. We may, however, state that light is really an electro-magnetic wave-motion, similar to that used in wireless-telegraphy, the wave-length in the case of light being enormously smaller than the wave-length of those waves used in wireless-telegraphy. The waves are tranverse waves (see p. 235), that is, the displacement takes place at right angles to the direction in which the wave is moving.

One of the characteristics of a wave-motion is the velocity with which the waves travel, and hence we proceed to give a short account of the methods by which the velocity of light has been measured.

Römer's method.

The first determination of the velocity of light was made by Römer, who measured the time light takes to traverse the earth's orbit. The principle of the method is this: Suppose a light signal is repeated at equal intervals of time t, at a station A, and that an observer at a station B observes the interval between these signals. Since the distance between A and B is constant, the observed interval will be t, although each flash will not be *seen* till some time after it has occurred, owing to the time taken for the light-waves to travel from A to B. Next suppose the observer at B to travel away from A, continuing to observe the intervals between the flashes as before. The intervals will now be greater than t, for each flash will now have to travel a greater distance before it reaches the observer than did the previous one, and the increase between successive intervals will be equal to the time light takes to traverse the distance through which the observer has moved in this interval. Hence the sum of the *increases* in the times of the intervals while the observer traverses a distance D will be the time light takes to traverse this distance. Hence if we know D and observe the intervals between the flashes we can calculate the velocity of light.

The signals used by Römer were the instants when one of Jupiter's moons is eclipsed owing to its passage into the shadow cast by Jupiter, and the distance over which the observer moved was the diameter of the earth's orbit.

The accuracy of the above method is limited by the accuracy with which the diameter of the earth's orbit is known, and more accurate measurements can be made if the velocity is measured between two points on the surface of the earth, in which case the distance can be measured directly. The first to perform this experiment was Fizeau, who in 1849 measured the velocity of light by using a method depending on the eclipsing of a source of light by the teeth of a rapidly rotating wheel, the principle of the experiment resembling Römer's method.

A bright source of light was placed at L (Fig. 217), and after passing

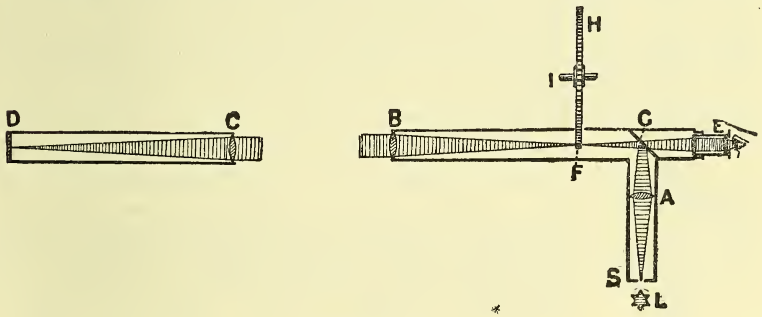


FIG. 217.

through a lens A, a certain proportion of the rays of light was reflected from the surface of an unsilvered plate of glass G, placed at an angle of 45°. The reflected rays came to a focus at F, this point being the principle focus of a second lens B. Thus the light left B in a parallel beam, which, after traversing a distance of about four miles, fell on a lens C, and was brought to a focus at the surface of a spherical mirror D. The curvature of this mirror is such that the lens C is at its centre of curvature, and hence the rays are reflected back along their path, so that on emerging from the lens C they again form a parallel beam. This reflected beam falls on the lens B, is brought to a focus at F, and then falls on the plate of glass G, where some of the rays will be reflected, and some will be transmitted and enter the eye-piece E, so that a bright star will be seen by the observer, formed by light which has travelled to D and back again. A toothed wheel, H, which can be rapidly rotated round an axle, I, is so arranged that when a tooth passes F the light is intercepted, but when a space passes F the light is allowed to pass.

Fizeau's method.

If the wheel is slowly rotated, an observer at E will see a bright star, when a space passes F, while when a tooth passes there will be darkness, so that as the wheel rotates the star alternately appears and disappears; but if the speed is such that more than twenty teeth pass per second, owing to the persistence of vision, a permanent star will be seen.

If light took no time to travel from F to D and back again, then all the light that passed through any space would be able to pass back again through the same space, since by supposition the light takes no time to travel from F to D and back, and hence the wheel would not have moved between the starting of the light and its arrival back at F. If, however, the light takes an appreciable time to travel, then, as the speed of the wheel is gradually increased, it will eventually rotate so fast that by the time the return light reaches F the wheel will have turned so that a tooth will have moved round, and will occupy part of the space which was occupied by a space when the light started, so that part of the return light will be cut off by the tooth, and hence the star seen at E will be of decreased brightness. As the speed is further increased, more and more of the returning light will be intercepted by the succeeding tooth, till finally all the light which gets through a space is, on its return, intercepted by the succeeding tooth, and no star is seen. If the speed is yet further increased, the returning light will begin to get through the space next after the one through which it passed on its way out, and a bright star will again be seen. Hence, as the speed is increased, the star alternately appears and vanishes.

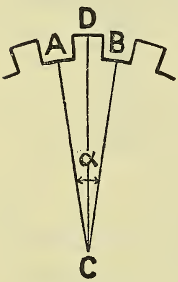


FIG. 218.

If the wheel contains d teeth, then the angle ACB (Fig. 218) subtended by the interval between two consecutive teeth or spaces at the centre will be $360°/d$ or $2\pi/d$ in circular measure. Hence half this angle, or π/d, is the angle through which the wheel has to turn so that a tooth may exactly occupy the position previously occupied by a space.

If the wheel makes n revolutions per second when the first eclipse occurs, the angle swept out by any radius AC in one second will be $2\pi n$. Hence the time taken to turn through the angle π/d will be $\frac{\pi}{d}\Big/2\pi n$, or $1/2dn$. If l is the distance between F and D, the distance passed over by the light while the wheel has turned through an angle π/d is $2l$. Hence the light has travelled a distance $2l$ in a time $1/2dn$, or the velocity of light V is given by

$$V = \frac{2l}{\frac{1}{2dn}} = 4ldn$$

In one of the experiments, l was equal to 8633 metres, the wheel had 720 teeth, and when the star was first eclipsed it made 12·6 revolutions per second, so that

$$V = 4 \times 8633 \times 720 \times 12{\cdot}6 = 313274000 \text{ metres per second.}$$

More recent experiments made by Cornu, using this method, gave 300,400 kilometres per second, or 186,662 miles per second, as the velocity of light.

Foucault's method.

In the year 1850, Foucault succeeded in measuring the time light took to travel over a distance of about twenty metres. His method consists in causing a beam of sunlight to fall on a slit S (Fig. 219), by means of a heliostat. The light transmitted by S passes through an unsilvered glass plate G, falls on a convex lens A, and then on a plane mirror B, which can be rapidly rotated round an axis perpendicular to the plane of the figure. For one position of the mirror B, the reflected light falls upon a second mirror C. This

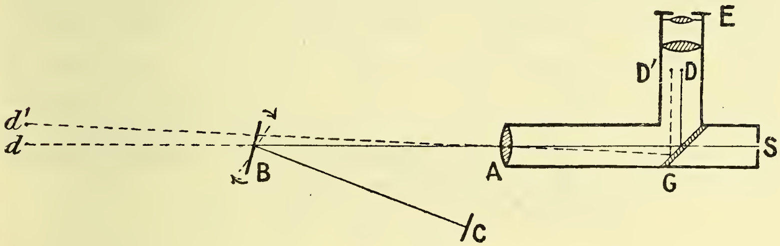


FIG. 219.

latter is a concave mirror, the radius of curvature being equal to BC. Hence, if the mirror B is at rest, the light reflected at C will retrace its path, being partly reflected at G, so as to form an image of the slit S at D, this image being observed through an eye-piece E. If the mirror B is rotated, the light is only reflected back from C once in each revolution, if the mirror is only silvered on one side. If, further, during the time taken for the light to travel from B to C and back to B, the mirror has appreciably turned, the return rays will be reflected by B in a slightly different direction to that which they would have taken had the mirror been at rest, and the image D will be displaced to D′, the amount of the displacement being read off on a scale placed at D. In order to count the speed of rotation of the mirror B, which was driven by a small steam-turbine, Foucault placed a toothed wheel so that the teeth were illuminated by the intermittent beam of light reflected from the rotating mirror. If the wheel was rotating at such a speed that during the interval between

two flashes one tooth had just moved into the position occupied at the previous flash by the preceding tooth, then the wheel would appear to be at rest. When this is the case, the time taken by the mirror to make one revolution is equal to the time taken by the wheel to turn through the angle ACB (Fig. 219), that is, $1/dn$, where d is the number of teeth in the wheel, and n the number of revolutions it makes in a second. Hence the number N of revolutions made by the mirror in one second is dn.

In order to calculate the velocity of light, we require to determine through what angle the mirror B must turn to shift the image formed at the eye-piece from D to D′. Since an image of the slit S is produced at C, we may if we please look upon the problem as to determine through what angle θ must B turn so that the image of C formed by the mirror and the lens A may shift from D to D′. Let the two positions of the mirror B be indicated by the full and dotted lines. Then for each position of the mirror B there will be an image of C formed in B. Let d and d' be these images; then, as shown in § 109, d and d' lie on the circumference of a circle having its centre on the axis about which the mirror turns, and passing through C. Further, the arc intercepted between d and d' subtends an angle 2θ at the centre of this circle. If, however, the distance BC is called l, we have $dd' = 2\theta l$. Now as far as the lens A is concerned, instead of taking C as the object and considering the two positions of the mirror, we may suppose the mirror B removed and take the two images d and d' as the objects. Then

$$\frac{dd'}{\mathrm{DD}'} = \frac{d\mathrm{A}}{\mathrm{AG} + \mathrm{GD}}$$

Using the letters a, b, c to represent the distances DD′, BA, AG + GD respectively, we have

$$\theta = \frac{dd'}{2l} = \frac{a(l+b)}{2lc}$$

Since the mirror makes N rotations per second, the time it takes to turn through an angle θ is $\theta/2\pi N$, and in this time the light has travelled from B to C and back, that is, a distance $2l$. Hence the velocity, V, of light is given by

$$V = \frac{4\pi Nl}{\theta} = \frac{8\pi Nl^2c}{a(l+b)}$$

Using this method, Michelson has obtained 299,853 kilometres per second as the velocity of light, with a possible error of ± 60 kilometres.

Knowing the velocity of light in air, if we know the refractive index

from air to any other medium then we can calculate the velocity of light in this medium. For let AB (Fig. 220) be the line of separation between air and water. Let PP′ represent a wave-front (p. 245) in the air, then if v_a is the velocity of light in air, the time taken for the point P′ on the wave-front to reach the second medium will be $P'O/v_a$. During this time the point P on the wave-front will have travelled into the water, and if v_w is the velocity in water, it will have travelled a distance $\frac{P'O}{v_a} \times v_w$. If then, with centre P and radius equal to this distance, we describe a circle, and then from O draw a tangent OR, OR will represent the wave-front in water at the instant when the point P′ on the wave-front PP′ reaches O (see § 92). If ON and PN′ are normals to AB, we have

Connection between refractive index and velocity of light.

$$_an_w = \frac{\sin NOP'}{\sin N'PR}$$

Now in the triangle PP′O the angle at P′ is a right angle, hence the two angles P′PO and P′OP are together equal to a right angle. But the angles

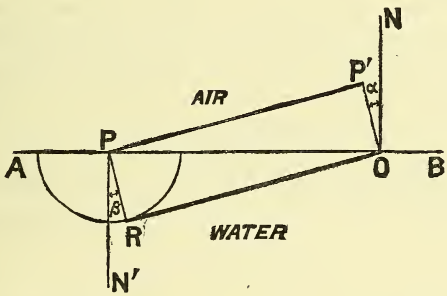


FIG. 220.

P′ON and P′OP are also together equal to a right angle. Hence the angle P′ON is equal to the angle P′PO. In the same way, the angle N′PR is equal to the angle POR. So that

$$_an_w = \frac{\sin NOP'}{\sin N'PR} = \frac{\sin P'PO}{\sin POR} = \frac{P'O}{PO} \div \frac{PR}{PO} = \frac{P'O}{PR}$$

$$= P'O \div \frac{P'O}{v_a} v_w = \frac{v_a}{v_w}$$

Or the ratio of the velocity of light in air to that in water is equal to the refractive index from air to water. It will thus be seen that light travels more slowly in water than in air, for the refractive index is 1·33.

If a light-wave traverses a distance d in a medium of which the refractive index is n, the time taken will be the same as that which would be occupied in traversing a distance nd in *air*. Hence nd is called the equivalent air-path to the distance d in the given medium.

Equivalent air-path.

This result can be applied to find the relation between the radii of curvature of the faces of a lens and the focal length. Taking the simple case of a plano-convex lens, AB (Fig. 221), suppose a parallel beam of light I_1A, I_2O incident on the plane face. Since the rays are parallel it follows that the wave-front is plane, and hence a given wave will reach A and O simultaneously. Consider two portions of the wave which traverses the lens, one portion which travels along OCF, and the other along AF. If F is the principal focus these two portions of the wave must reach F at the same instant, otherwise they would interfere (see p. 244). Thus the wave that travels from A to F in air must occupy the same time as the wave that travels from O to C in glass and from C to F in air. That is, the equivalent paths must be the same. Hence

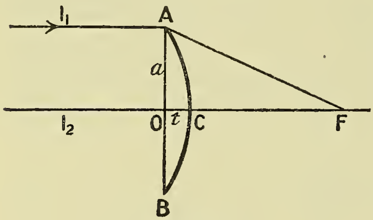


FIG. 221.

$$AF = OC \, . \, n + CF \quad . \quad . \quad . \quad . \quad (a)$$

If a is the semi-diameter AO of the lens, t the thickness, OC, and r the radius of curvature of the surface ACB, we have (segments of chords of a circle)

$$a^2 = (2r - t)t = 2rt$$

if the thickness of the lens is small compared to the radius of curvature.

Also, if f is the focal length, OF, of the lens,

$$AF = \sqrt{(a^2 + f^2)} = f\sqrt{\left(1 + \frac{a^2}{f^2}\right)} = f\left(1 + \frac{a^2}{2f^2}\right)$$

if f is large compared to a. Hence from (a)

$$f\left(1 + \frac{a^2}{2f^2}\right) = nt + f - t$$

or

$$\frac{1}{f} = (n - 1)\frac{1}{r} \quad . \quad . \quad . \quad . \quad (132)$$

a result which agrees with the formula given on p. 315, for with a plano lens $r_2 = \infty$ and hence $\frac{1}{r_2} = 0$.

120. Interference.—One of the chief characteristics of a wave-motion is that under suitable conditions two systems of waves may interfere, that is at some places they may destroy each other while at others they strengthen each other. In the case of the ripples on the surface of mercury (p. 242) we obtained interference between the waves starting from two separate centres of disturbance, though both styles which dip in the mercury were attached to the same fork, so that the waves always started in the same phase. Interference can also be obtained when the styles are attached to *different* forks if care is taken to tune these to correct unison, in which case also the phase relation between the forks remains constant.

In the case of light, all attempts to obtain interference between light-waves emitted by two neighbouring sources, or even two separate portions of the same source, fails, this failure being due to the fact that the phase of the light vibrations given out by a source suffers rapid and abrupt changes, so that in the case of two separate sources the phase of the emitted light may be the same for, say, a thousand vibrations, a crest leaving each simultaneously, and thus producing darkness at a certain point P; then suddenly the phase of the light given by one source will change, so that while a crest is leaving one source a trough will be leaving the other, and thus the waves now strengthen each other at P. Since such changes, if they occurred, say, a hundred times a second, would not be visible owing to persistence of vision, and during a hundredth of a second 5×10^{12} vibrations of yellow light take place, it is evident that it is not necessary to assume that the changes of phase take place so very frequently, in order to explain the absence of interference between the light from two independent sources.

If instead of two separate sources we take as sources two images of the *same portion* of a luminous body, then, whenever a change in phase takes place in the source, the corresponding change in phase will take place *simultaneously* at the two images; and hence if interference is produced at a given point before the change of phase, it will also be produced after.

Fresnel's mirrors.

Fresnel, in order to demonstrate the interference of light, devised two arrangements for producing two image sources in such positions that they can produce interference. One of his arrangements consists of two mirrors, AB and BC (Fig. 222), inclined at an angle of very nearly 180°, so that a luminous point at P will produce two images, one at P′ by reflection in the mirror AB, and the other at P″ by reflection in the mirror BC. In § 109 we have proved that P′, P″ both lie on a circle of which B is the centre. Hence if we join P′P″, bisect this line at F, and join FB, FB will be at right angles to P′P″. If FB is produced to meet a screen on which the reflected light is received at O, then the point O is equidistant from P′ and P″.

As far as the reflected light is concerned, we may regard it as coming from the images P′ and P″, so that the length of the path of any ray which, leaving P, is reflected at one of these mirrors and strikes a screen DE is the same as if it came from P′ or P″, as the case may be. At the point O of the screen, which is equidistant from the images, the light-waves will assist one another, since they always leave P′ and P″ in the same phase, these points being images of the same source.

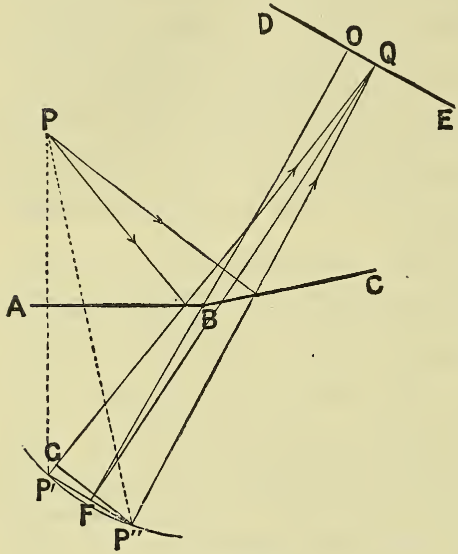


FIG. 222.

There will be interference at a point such as Q, if the difference between the distances P′Q and P″Q is equal to half a wave-length, for then the vibrations from P′ will reach Q in the opposite phase to those from P″.

The manner in which the waves coming from P′ and P″ at some points on the screen strengthen each other, and at other points annul each other, is made clear by the diagrammatic representation given in Fig. 223. In this figure the waved lines represent the displacements proceeding from the two sources P′, P″ along the lines P′O, P″O, P′Q_1, P″Q_1, &c.; and it will be seen that the displacements produced at the points O and Q_2, due to the two sets of waves, are in the same phase, so that the resultant displacement is twice the displacement due to either. It will also be seen that there are an equal number of waves between P′ and O and P″ and O, but that there is one more wave between P′ and Q_2 than between P″ and Q_2. At the point Q_1 the waves from the two sources are in opposite phase and destroy each other, and here there is half a wave more between P′ and Q_1 than between P″ and Q_1.

If with centre Q (Fig. 222) and radius QP″ we describe an arc of a circle cutting QP′ in G, then GP′ will be the difference between the paths P′Q and P″Q.

Since in practice the distance between the images P′ and P″ is excessively small compared with the distance of either of them from the screen, the arc P″G may be taken as a straight line, which is practically perpendicular to QP′ and QF.

Since P′P″ is perpendicular to OF, and GP″ is perpendicular to QF, the angle OFQ is equal to the angle GP″P′. Hence calling this angle A and the distance between the images $2d$, we have—

$$GP' = 2d \sin A = 2d \cdot \frac{OQ}{QF}$$

or since QF is practically equal to OF, A being very small,

$$GP' = 2d \cdot \frac{OQ}{OF}.$$

Calling the distance BF between the images and the mirrors p, that

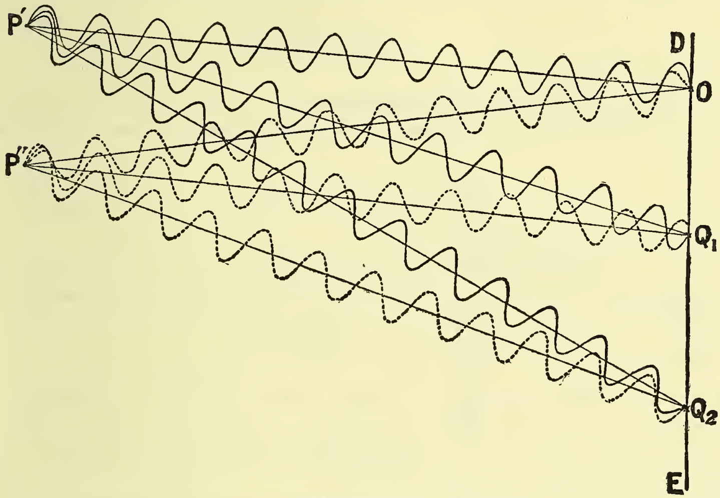


Fig. 223.

between the mirrors and the screen q, and the distance of the point Q from O x, we get, since OF $= p + q$,

$$GP' = \frac{2dx}{p+q}$$

Now if l is the wave-length of the light given out by the source P, interference will be produced at Q whenever the difference between the paths P′Q, P″Q is such that the light reflected from the two mirrors arrives at Q in opposite phases. This will occur when the paths differ by any odd number of half wave-lengths, for the ether at points separated by half a wave-length or any odd number of half wave-lengths is vibrating in opposite phases.

Hence if GP′ is equal to $(2y+1)\frac{l}{2}$, where y is any whole number, we shall get interference at Q.

Hence if there is interference,

$$\frac{2dx}{p+q}=(2y+1)\frac{l}{2}$$

or

$$x=(2y+1)\frac{l}{2}\cdot\frac{p+q}{2d} \quad . \quad . \quad (133)$$

If the difference in path, GP′, is equal to an even number of half wave-lengths, *i.e.* to a whole number of wave-lengths, the light will reach Q in the same phase, and hence a bright band will be produced at Q. When this occurs,

$$x=2y\cdot\frac{l}{2}\cdot\frac{p+q}{2d} \quad . \quad . \quad . \quad (134)$$

In this experiment the two image sources P′, P″ play the same part as the two needle points attached to the tuning-fork in the interference

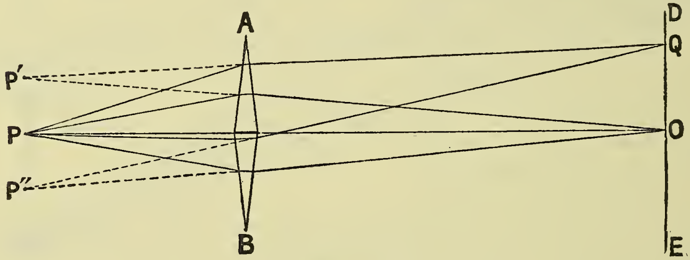


Fig. 224.

experiment with capillary waves on the surface of mercury (§ 91). If a line in Fig. 157 were drawn parallel to the line joining the two centres of disturbance, this would represent the screen in the optical experiment, and wherever this line cuts one of the interference curves in the figure would correspond to a dark band in the optical experiment, while half-way between each curve the mercury surface is disturbed by the combined action of the two centres of disturbance, and this corresponds to a bright band in light.

In the second method used by Fresnel, the light from a luminous point P passes through two narrow-angle prisms, AB (Fig. 224), placed with their bases in contact, forming what is called Fresnel's bi-prism. After passing through the bi-prism, the light travels as if it came from the two points P′ and P″, and interference is produced on a screen placed at DE, as in the previous case. Calling, as before, the distance of the images from the bi-prism p, that of the screen from the bi-prism q, and the distance between the images

Fresnel's bi-prism.

$2d$, then a dark band will be produced by interference at Q, if the distance OQ or x is such that

$$x=(2y+1)\frac{l}{2}\frac{p+q}{2d} \quad . \quad . \quad . \quad . \quad (135)$$

as in the case of the two mirrors.

Fresnel's experiments may be used to determine the wave-length of the light used, for by measuring x we can calculate l. It is, however, generally more convenient to measure the distance between the dark interference bands than to measure the distance of a dark band from the central bright band. The distances of the first few dark bands from O are given by

$$x_1=\frac{l}{2}\cdot\frac{p+q}{2d}$$

$$x_2=\frac{3l}{2}\frac{p+q}{2d}$$

$$x_3=\frac{5l}{2}\frac{p+q}{2d} \text{ \&c. \&c.,}$$

so that the distance (z) between two consecutive bands is

$$z=l\frac{p+q}{2d} \quad . \quad . \quad . \quad . \quad . \quad (136)$$

In this expression $\frac{p+q}{2d}$ is independent of the wave-length of the light used, so that we see that the greater the wave-length l, the further apart are the interference bands. It is found by experiment that with red light the bands are further apart than with violet light, so that the wave-length of violet light must be less than that of red light.

If white light is used, the violet light will be destroyed nearer to the centre O than the red light, so that this red left over will produce a red band on either side of the central bright band, which will be white, for the light of all wave-lengths arrives in the same phase at O. A little further out from O the red light will be destroyed by interference leaving the violet light, so that here a violet band will be produced. Hence when white light is used the first dark band will be bordered with red on the inside and violet on the outside. The distance between the points where the red and violet are destroyed will increase with each successive band, until finally there will be overlapping between one bright band for the violet and the previous bright band for the red; and at some distance from O, the overlapping will be so considerable that white light will be reproduced, and so no bands will be discernible.

Since only a few interference bands are obtained with Fresnel's

arrangements, they do not form a very convenient method of measuring the wave-length of light. A much simpler and more accurate method is to use a diffraction grating which consists of a number of equidistant parallel lines ruled on a plate of glass, or of speculum metal. In order to explain the action of a grating, we shall suppose that it consists of a series of equally spaced opaque lines ruled on a plate of glass, the width of each line being equal to the space between two adjacent lines.

The diffraction grating.

Let AB and CD (Fig. 225) be two adjacent spaces, and suppose a beam of parallel light to be incident on the grating normally, *i.e.* parallel to NA, so that the incident wave-fronts are parallel to the grating. We may then look upon each point in the spaces AB, CD, &c., as a centre of disturbance from which light-waves are propagated, all these waves starting in the same phase. Consider two of these centres of disturbance, one at A and the other at C. The disturbances from these centres will reach all points at equal distances from A and B in the same phase, and so will strengthen one another.

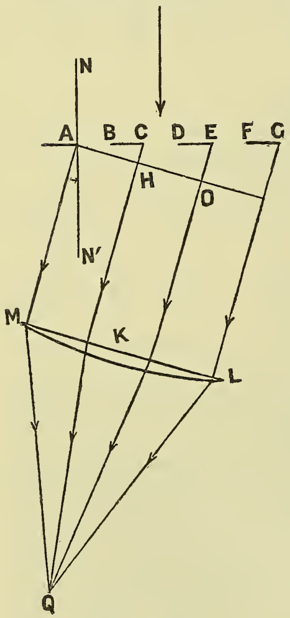


Fig. 225.

At any other point Q, however, the disturbances need not be in the same phase. If, as is always the case, Q is at very great distance from the grating compared to AB, or, as shown in the figure, a lens L is interposed to form an image at its principal focus, we may take the lines AM and CK as parallel, and both inclined to the normal to the grating at an angle A.

From A draw AH perpendicular to CK or AM: then, since N′A is perpendicular to CA, and AH is perpendicular to AM, the angle CAH included between CA and AH is equal to the angle A included between N′A and AM. Therefore

$$CH = AC \sin A,$$

or if d is the combined width of a space and a line, so that AC $= d$,

$$CH = d \sin A.$$

Now the waves starting from A and C will be in the same phase when they reach Q, and therefore will strengthen each other, if the difference in the paths AM and CK is equal to an even number of half

wave-lengths.[1] Hence the condition for the production of a bright band at Q is that

$$d \sin A = 2y\frac{l}{2}$$

or

$$\sin A = \frac{yl}{d} \quad . \quad . \quad . \quad . \quad (137)$$

If CH is equal to an odd number of half wave-lengths, interference will be produced at Q, the condition for a dark band being

$$d \sin A = (2y+1)\frac{l}{2}$$

What we have said with regard to the two centres A and C will also apply to each pair of centres taken in AB and CD, so that the above equations also give the conditions for the production of a bright or dark band at Q, when the whole of the two spaces AB and CD are operative. A similar argument holds with regard to the next two spaces, and so on, so that the above equations apply to the grating taken as a whole.

If we consider the first bright band, then, as d is constant and $\sin A = l/d$, it is evident that the value of A will vary with the wave-length of the light, so that by measuring the angle A for the first bright band produced by different coloured lights, we can calculate the wave-length of these lights. If white light is used, the positions of the bright bands will be different for the different colours, and hence a spectrum will be produced.

When using this method to measure the wave-length of light the grating is mounted on the table of the spectrometer (Fig. 193), with its surface normal to the light coming through the collimator and the rulings on the grating parallel to the slit. The telescope is then turned to view the bright bands on either side of the central bright band corresponding to the undeviated light, and the difference between the readings gives 2A.

If white light is used, a series of spectra will be obtained corresponding to the cases where y is made 1, 2, 3, &c., in equation 137, called spectra of the first, second, &c. order.

The dispersion obtained with a grating increases as d decreases, that is as the number of lines per unit length increases. The dispersion in the *y.th* order spectrum is y times that in the first order spectrum. Since, however, the spectra of higher orders overlap and are also often faint, it is generally inconvenient to use any but the first or second. Michelson

[1] After striking the lens the waves will be brought to a focus at Q, and the virtual length of the paths depends on the constants of the lens. The virtual length, that is, the length allowing for the fact that light travels more slowly in glass than in air, of all the paths from the first surface of the lens to the focus Q is the same, so that any difference of phase which exists at M and K will persist when the waves reach Q.

has devised an arrangement, called an echelon grating, in which *bright* spectra of a very high order can be observed. With this instrument, however, an auxiliary spectroscope must be employed so that only light of very nearly a single wave-length is admitted, and hence overlapping is prevented.

121. Colours of Thin Films. Newton's Rings.—The bright colours shown by thin films of transparent substances, such as soap-bubbles, films of oil on the surface of water and the like, are due to interference.

Suppose a beam of light is incident very nearly normally on a soap-film, ABCD (Fig. 226), along the line IM. Some of the light will be reflected along MR, while part will penetrate the film, and this will be partly reflected along NM′R′ and part, transmitted along NT. The waves which have been reflected at M and those which have been reflected at N may interfere, and we now proceed to find what must be the thickness of the film for such interference to take place. The difference in path is equal to twice the thickness, T, of the film. Now, as we have seen (see p. 345), the velocity of light in soap solution will be n times that in air, where n is the refractive index of the solution. Hence the waves will take as long to traverse the path 2T in the film as they would to traverse a distance 2Tn in air. Thus the part of the wave which has been reflected at M will have travelled through a distance 2Tn while the other part has been traversing the film twice.

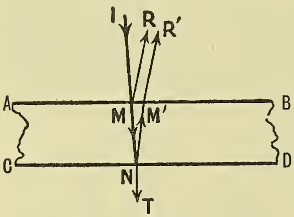


FIG. 226.

If the thickness of the film is *excessively* small, 2Tn will be very small, and hence according to the above the two sets of waves ought to be in the same phase and hence strengthen each other. Observation, however, shows that if the film is excessively thin interference takes place, so that the film looks black. This discrepancy is due to the fact that the reflection at M and N takes place under different conditions. Thus at M reflection occurs as the light passes from a less dense medium (air) to a more dense one (solution), and in such a case there is reflection with a change of sign. This resembles reflection at the closed end of an organ-pipe, or at the fixed end of a string. At N, on the other hand, reflection takes place at a surface separating a denser from a less dense medium, and there is no change of phase. This corresponds to reflection at the open end of an organ-pipe or at the free end of a string. Hence since one of the interfering wave-trains suffers a change of phase, equivalent to an increase of path of half a wave-length, and the other does not, they will annul each other's effect when the extra path due to the film

Interference in reflected light.

is equal to a whole wave-length or any integral number of whole wave-lengths, that is when

$$yl = 2Tn \quad . \quad . \quad . \quad . \quad (137)$$

where y is any whole number.

If the thickness of the film varies and the incident light is white a series of coloured patches will be produced, for light of different colours, *i.e.* different wave-lengths, will be destroyed by interference at different points, according to the thickness, and by the loss of such constituent the remaining light will appear coloured.

In addition to the interference between the reflected beams considered above, interference may also take place in the light which is transmitted through the film. The paths of the wave-trains which interfere in this

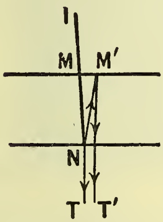


FIG. 227.

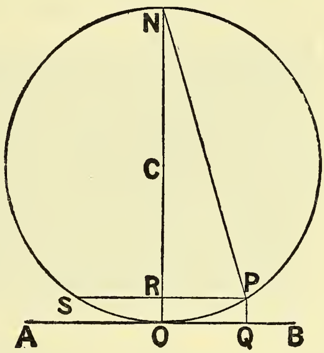


FIG. 228.

way are shown in Fig. 227. One train passes straight through the film, while the other is twice reflected within the film. The additional path is $2Tn$ as before, but since the reflection at both N and M′ is at a surface separating a more dense from a less dense medium, this train loses two half wave-lengths, and hence no allowance on this account has to be made. The transmitted beams therefore interfere, producing darkness, when $2Tn$ is equal to an odd number of half wave-lengths, *i.e.* when

Interference in transmitted light.

$$2Tn = (2y-1)\frac{l}{2} \quad . \quad . \quad . \quad . \quad (138)$$

If a convex lens of large radius of curvature is pressed against a flat glass plate an air-film will be included, of which the thickness is zero at the point of contact, but increases as we go out from this point. Interference will take place in this film, and if we are using monochromatic light, wherever the thickness is given by the expression (137) above a dark ring will be produced in the reflected light, while where the thickness is given

Newton's rings.

by (138) a dark ring will be produced in the transmitted light. These rings are known as *Newton's rings*. If white light is used there will be fewer rings observed, and these will be coloured.

If SOP (Fig. 228) is a section of the convex surface, the thickness of the air-film at PQ is given by

$$SR\,.\,RP = PQ(NO - PQ),$$

for PQ = RO and SR, RP are the segments of one chord of the circle SOPN, and RO and NR are the segments of another intersecting chord. Hence if the radius of the ring, RP, is r_y and the radius of curvature of the lens is R, we have

$$r^2_y = PQ(2R - PQ)$$

or since PQ is always small when interference takes place, so that $\overline{PQ}^2$ can be neglected,

$$PQ = \frac{r_y^2}{2R}$$

Hence if r_y is the radius of the *y.th* dark ring of the *reflected* system, since $n = 1$ for air

$$yl = \frac{r_y^2}{R}$$

or

$$r_y^2 = ylR \quad . \quad . \quad . \quad . \quad (139)$$

So that the squares of the radii of the rings are as the natural numbers 1, 2, 3, &c.

In the case of the transmitted light, in the same way the radius of the *y.th* dark ring is given by

$$r_y^2 = \frac{(2y+1)}{2}lR \quad . \quad . \quad . \quad (140)$$

It will be observed that the radii of the dark rings of the transmitted system are equal to the radii of the bright rings of the reflected system, and *vice versâ*.

122. Diffraction.—In an earlier section we referred to the rectilinear propagation of light, and the formation of shadows as a characteristic of light. Now if light is a wave-motion the question arises why is it that light proceeds in straight lines in this way while sound-waves certainly do not. Thus if we have a hole in a screen and place a source of light on one side, the only part of the room illuminated on the far side of the screen is that opposite the hole. If, however, a sounding body is placed in the position previously occupied by the light, the sound can be heard at all parts of the room on the far side of the screen. The full explanation of this difference would be beyond the scope of this book, but we may give a hint of the solution by showing that (*a*) sound shadows can be produced, and (*b*) that light does to a certain

extent turn round the corner of a screen; the difference between the behaviour of sound and light waves being due to the fact that the obstacles used in the case of sound are generally small compared to the wave-length, while in the case of light the opposite is the case.

If a very high note is sounded, of which the wave-length is not more than a centimetre, then well-marked shadows can be detected. Thus if a shrill whistle is sounded and a drawing-board is set up as an obstacle, the presence of a comparatively well-marked shadow can be shown by moving a sensitive flame (§ 100) about behind the board.

On the other hand, if a narrow slit is placed before a source of light and at some distance a sharp edge is set up, the edge being parallel to the slit, it will be found that if a screen is placed a little way behind the

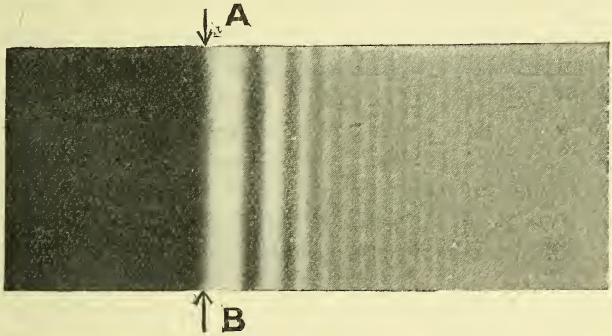


FIG. 229.

edge, the edge of the shadow is not quite sharp, but that there are a number of alternate bright and dark bands outside the geometrical shadow, while the light shades off gradually within the geometrical shadow. In Fig. 229 is shown a photograph of the edge of the shadow formed by a razor-edge as described above, the line where a series of straight lines from the slit to the edge would intersect the photographic plate would join the two arrows A and B. The bands produced are due to interference, the spacing of the bands depending on the wave-length of the light. It will be observed that the light at the bright bands just outside the edge of the shadow is brighter than at some distance away, *i.e.* that owing to the presence of the shadow the illumination at some points is greater than would be the case if no obstacle were present. The explanation of this phenomena will be found in Watson's *Text-Book of Physics*, § 383, and attention may be drawn to Fig. 373 there.

CHAPTER VI

PHOTOMETRY

123. Light Units.—We have now to consider the question as to the quantitative estimate of light, that is, study the relative brightness of different lights. Suppose we have a very small source of light which is emitting light uniformly in all directions. Then we may look upon the light as flowing out in all directions along straight lines radiating from the source. If a sphere be described round the source as a centre all the light emitted by the source will fall on the interior surface of this sphere, and as the flow of light is uniform in all directions the quantity of light which falls on each unit area of the surface of the sphere is the same, and is proportional to the total quantity of light emitted by the source. Hence we may take the quantity of light which falls on unit area of a sphere of unit radius as a measure of the quantity of light emitted by a source. Thus the light received by unit area at unit distance from a source is taken as a measure of the *illuminating power* of the source.

Illuminating power and intensity of illumination.

Since the total area of a sphere of unit radius is 4π, and on each unit of area of such a sphere described about a source of which the illuminating power is I, there falls a quantity of light I, the total quantity, or flux, of light which is emitted by the source is $4\pi I$.

The quantity of light which falls on unit area of a given surface, which is placed normally to the incident light, is called the *intensity of illumination* of the surface.

If the given surface consists of a sphere of radius d described with a source of illuminating power I as centre, and the source radiates light uniformly in all directions, the total quantity of light which falls on the surface is $4\pi I$, and the area of the surface is $4\pi d^2$. Hence the intensity of illumination is I/d^2. Thus the intensity of illumination produced by a small source varies inversely as the square of the distance. This law does not hold in the case of sources of light of large area except at distances which are great compared to the size of the source.

In the above we have spoken about the quantity of light emitted by a source or received by a screen, but have not said anything about the unit in which it is measured.

It must be remembered that we are here dealing with light with

reference to its effect on the *eye*, that is with the sensation produced by a source of light, and thus as we have no direct way of measuring this sensation we are obliged to fall back upon the device of arbitrarily selecting some source of light and taking its illuminating power as unity.

Unit of illuminating power. The candle.

The standard unit of illuminating power ordinarily adopted is that of a "standard candle." The standard candle is a sperm candle, 3/8ths of an inch in diameter, which burns 120 grains (7·776 grams) of wax per hour. This standard is a most unsatisfactory one, since it is very variable. For this reason other standards have been adopted, but they have been compared with a large number of standard candles, and their illuminating power is expressed as so many candle-power. Thus the Harcourt pentane standard lamp is adjusted to give an illumination equal to that which on the average would be produced by ten standard candles, and hence is called a ten candle-power standard. The three most important of these subsidiary standards are—(1) the Harcourt lamp referred to above, which burns pentane, and when in use has the flame adjusted to specified dimensions; (2) the Hefner lamp, which burns amyl acetate, and again has the flame adjusted to fixed dimensions; (3) the carcel, which is a lamp burning colza oil, and is not very much more constant than a standard candle.

Electric filament lamps are often used as subsidiary standards of illuminating power, their candle-power when a definite current is passing being determined by comparison with one of the standard lamps described above. Although the candle-power of an electric lamp when the electrical conditions are kept fixed is for a limited time much more constant than is that of any other source, yet there is a gradual change which precludes their adoption as fundamental standards.

If a source of which the illuminating power is one candle-power gives out light uniformly in all directions, the total quantity of light emitted is 4π, and this is taken as the unit of quantity or flux of light, and is called a *lumen*.

Units of intensity of illumination.

The intensity of illumination at a distance of one metre from the above source of one candle-power is the unit of intensity of illumination, and is called a *candle-metre or lux*. Thus in the case of a small source of I candle-power the total flux of light is $4\pi I$ lumens, and the intensity of illumination at a distance of d metres is $1/d^2$ candle-metres.

In the British system of units the unit intensity of illumination corresponds to that at a distance of a *foot* from a source of one candle-power, and this unit is called a candle-foot. Since one foot is equal to ·305 metres, the intensity of illumination at a distance of one foot from

a source of one candle-power is $(1/{\cdot}305)^2$ lux. Hence a candle-foot is equal to 10·75 lux.

If the area of the luminous surface of a source of light is a, and the candle-power is I, the quotient I/a is called the *intrinsic brightness* of the source.

124. Photometry.—Although the eye is only capable of very roughly estimating the relative intensity of the illumination of two surfaces, yet it is capable of telling with considerable accuracy when two adjacent surfaces are illuminated with equal intensity. Hence, to compare the illuminating powers of two sources, they are so arranged that two adjacent patches on a screen, each patch being illuminated by only one of the sources, appear illuminated with the same intensity. Now if I_1 and I_2 are the illuminating powers of the sources, and d_1, d_2 the distances between each source and the patch of the screen which *it illuminates* when the intensity of illumination of the two patches is the same, we have that the intensity of the illumination due to one source at its patch is $I_1/d_1{}^2$, and that due to the other at its patch is $I_2/d_2{}^2$. Hence, since these intensities are equal, we have

$$\frac{I_1}{d_1{}^2}=\frac{I_2}{d_2{}^2}$$

or

$$\frac{I_1}{I_2}=\frac{d_1{}^2}{d_2{}^2}$$

or the illuminating powers of the two sources are directly as the square of the distances at which they respectively produce equal intensities of illumination.

There are a number of arrangements in use for obtaining the two patches, illuminated each by a separate source, in such relative positions that any difference in their intensities of illumination may most easily be detected by the eye. One of the simplest *photometers*, as the instruments used for comparing the illuminating powers of two lights are called, is that due to Rumford.

A Rumford's photometer consists of an upright screen AB (Fig. 230), which is covered with white unglazed paper—white blotting-paper does very well—and in front of which is placed an upright opaque rod R about an inch in diameter. There are two scales, M and N, inclined at the same angle to the screen, along which the two sources, P and Q, can be moved. If the source P only were present, the rod R would cast a shadow *fd* on the screen, while if the light Q only were present, the shadow of the rod would be at *af*. Hence when both lights are present, while the parts A*a* and *d*B of the screen are illuminated by the two sources, the part *fd*, which receives no light from P owing to the inter-

position of the opaque rod, is only illuminated by Q, and the part *af* is only illuminated by P. If, then, the distances of the lights from the screen are adjusted till the two patches *fd* and *af* are equally illuminated, we shall have that the source P produces, at the distance of P from *af*, the same intensity of illumination as the source Q produces at the distance of Q from *fd*. Hence if d_1 is the distance of P from the part of the *screen* which it only illuminates, namely *af*, and d_2 is the distance of Q from the part of the screen it only illuminates, namely *fd*, we get, if I_1 and I_2 are the illuminating powers of P and Q respectively, that

$$\frac{I_1}{I_2}=\frac{d_1^2}{d_2^2} \quad . \quad . \quad . \quad . \quad (141)$$

The distance of the rod R from the screen is adjusted so that the shadows of the rod cast by the two lights just touch one another, as it is found that the eye can best judge when they are equally illuminated under these conditions.

Another form of photometer which is frequently used is Bunsen's

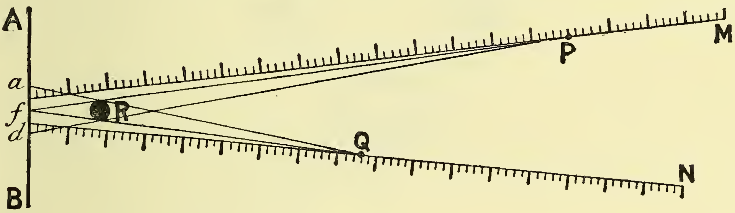


Fig. 230.

grease-spot photometer. This photometer consists of a small screen which has a central spot of grease, or in some other way is constructed so that the central portion is more translucent than the surrounding parts. If such a screen is held between the eye and a source of light, more of the light passes through the grease-spot than through the surrounding more opaque parts of the screen, so that the spot appears brighter than the surrounding paper. If, however, the screen is held against a dark background and illuminated from the front, the grease-spot will appear dark, for more of the light which is incident on the screen is transmitted through the spot than through the rest of the screen, and hence less is reflected or diffused so as to reach the eye by the spot than by the surrounding parts. If the screen is equally illuminated on both sides, then the spot diffuses less of the light received from the one source than the surrounding parts, but it transmits more of the light from the other source, so that these two effects just neutralise one another, and the spot appears of the same brightness as that of the surrounding paper.

The screen with the grease-spot is placed between the two sources whose intensities have to be compared, and moved about till the grease-spot can no longer be distinguished from the rest of the screen. If, when this adjustment has been made, the distances of the two sources from the screen are d_1 and d_2, we have, as before,

$$\frac{I_1}{I_2} = \frac{d_1^2}{d_2^2}$$

In using Bunsen's photometer, it is of assistance if both sides of the screen can be seen simultaneously. The usual arrangement employed to secure this end is a system of two mirrors inclined at 45° to the screen. With this arrangement one side of the screen is seen

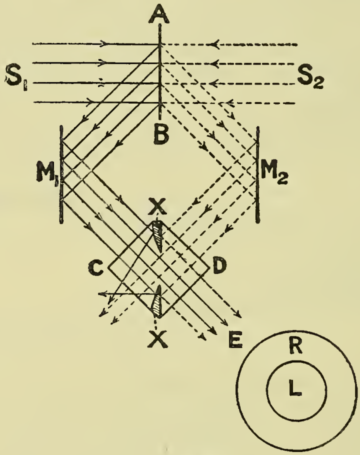


FIG. 231.

with one eye and the other with the other, so that if, as is generally found to be the case, one eye is more sensitive than the other, a wrong setting may be made. This source of error is removed in the Lummer-Brodhun photometer, which consists of an opaque screen AB (Fig. 231), each side being illuminated by one of the sources which are to be compared. The two sides of the screen are viewed by means of two plane mirrors, M_1 and M_2, and a double glass prism, CD. This prism consists of two right-angled prisms, the longest face of one being partly bevelled away, fastened together with Canada balsam. Owing to total internal reflection (§ 111), the central rays reaching an eye at E come from the left-hand face of the screen, and the surrounding rays come from the right-hand side, as is shown in the figure. Hence the observer moves the photometer till the central patch L and circumferential parts R appear of the same brightness, when the intensity

of the illumination on the two sides of the screen is the same; the relative powers of the two sources are then as the squares of their distances from the screen.

If the colour of the light emitted by the two sources to be compared is different, the comparison of their illuminating power is a matter of great difficulty. Further, although all observers will agree as to the relative candle-power of two sources when the character of the light given out by the two is the same, this is not the case when the lights differ in colour.

CHAPTER VII

RADIATION

125. Temperature Radiation.—In previous sections we have referred to the spectrum produced when white light is refracted through a prism, as appearing to the eye to extend from red to violet. If, however, we employ other methods of exploring the spectrum, it is found to extend far beyond the visible portion at both ends. Thus if a delicate thermometer, or preferably a thermopile, is made to traverse the spectrum produced by passing the light from an electric arc through a prism, it will be found that not only does the heat received by the thermopile gradually increase as we go from the violet end of the spectrum to the red, but that this increase continues for some distance beyond the visible spectrum. This extension of the spectrum beyond the red is called the *infra-red*, and corresponds to radiation having a wave-length *greater* than that of any light. We here use the term radiation to include not only the wave-motion which is of such wave-length as to be perceptible as light, but also all wave-motions of a similar kind, whatever their wave-length.

Spectrum extends beyond the visible portion.

In exactly similar manner the spectrum is found to extend beyond the limits of the visible violet, this portion being called the *ultra-violet*, and corresponding to radiation of *smaller* wave-length than any in the visible portion of the spectrum.

The infra-red portion of the spectrum corresponds to what is generally called radiant heat. Thus it is to these waves that we owe most of the sensation of warmth we experience when the hand is placed a little distance from a kettle containing boiling water. The ultra-violet radiation plays the most important part in photography, since it is to waves of these small wave-lengths that the chemical changes which take place in the photographic plate are chiefly due.

If a definite portion of the spectrum, and by the term spectrum we include both the visible and invisible parts, is allowed to fall on an absorbing surface, we can from the rise in temperature of this surface obtain a measure of the rate at which energy is being received, *i.e.* we can measure the energy corresponding to the given portion of the spectrum, and hence plot a curve showing how the energy varies throughout the spectrum obtained with the radiation from different sources.

In Fig. 232 are shown the energy curves corresponding to the spectra of the radiation from the sun and from blackened copper at different temperatures. It will be observed that as the temperature of the source of the radiation increases the spectrum extends further and further towards the shorter wave-lengths. Further, that as the temperature rises the energy at any particular wave-length increases, but in such a way that the wave-length, l_m, corresponding to the maximum is smaller the higher the temperature. The depressions in the energy curve for the

Effect of temperature of source on energy curve.

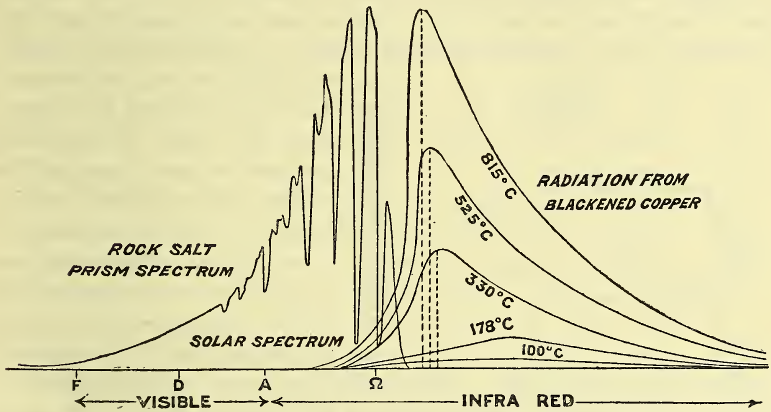


FIG. 232.

solar spectrum, indicating as they do the partial absence of radiation of certain wave-lengths, are due to absorption caused by the earth's atmosphere.

The total energy radiated by a source which is a full radiator,[1] that is the area of the energy curve of the spectrum, is proportional to T^4 where T is the temperature of the source on the absolute scale. This result is known as Stefan's law. The wave-length l_m corresponding to the maximum of the energy curve is connected with the temperature by the relation

Stefan's law and Wien's law.

$$l_m T = \text{constant} \quad . \quad . \quad . \quad . \quad (142)$$

This is known as Wien's law.

The laws of radiation given above have received an important practical application in their use in pyrometers for measuring tem-

[1] By a full radiator is meant a body the surface of which will entirely absorb all radiation which falls upon it. Since such a body would absorb all light it would look perfectly black at ordinary temperatures, and hence it is often called a perfectly black body. Lamp-black is very nearly a full radiator.

peratures too high to be measured in the ordinary manner, such as the temperatures of furnaces and the like.

The Fery radiation pyrometer consists of a concave mirror B (Fig. 233) which forms an image of the hot body, say a hole A in the side of a furnace, on a small blackened disc forming one junction of a thermocouple (§ 176) C. The radiation falling on the thermocouple is absorbed, and hence the temperature of the junction is raised so that a current is caused to pass through a galvanometer E connected to the couple by means of wires attached to the binding screws D. The instrument is directed so as to point directly at the opening in the furnace by looking through the telescope F, and arranging that the image of A formed by the mirror completely covers the disc C.

Radiation pyrometor.

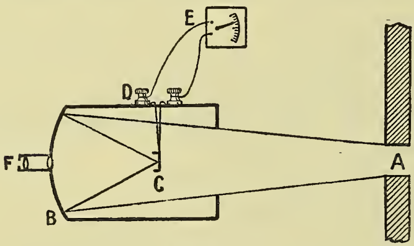


Fig. 233.

As long as the image is larger than the size of the disc attached to the junction, the reading is independent of the distance of the instrument from the hot body. The reading on the galvanometer is proportional to the energy received by the instrument, that is to the heat radiated by the hot body. Hence if the hot body is a full radiator the galvanometer deflection will, by Stefan's law, be proportional to T^4 where T is the absolute temperature of the hot body. Even if the hot body is not a full radiator it can be shown that if the radiation emerges through a small hole in the wall of a region in which the temperature is *constant* throughout, the amount of radiation is the same as would be the case with a full radiator at the same temperature. Thus the reading obtained through a hole in the side of a furnace will give the temperature even though the inside of the furnace is not a full radiator, as long as the temperature is fairly uniform throughout the inside.

A distinction must be drawn between the energy corresponding to any part of the spectrum and the light. Thus the maximum *brightness* of the solar spectrum is in the yellow, a little on the green side of the D line, while the maximum of the energy is, as shown in Fig. 232, in the infra-red. Hence the energy measurements even in the visible spectrum cannot be taken as a direct measure of the *light* given by a given source. Nevertheless, since it is obvious that it is only the energy of the radiation corresponding to the *visible* part of the spectrum which is of any value as far as the light-giving property of a source of light is concerned, we may take the ratio of the area of that part of the energy curve corresponding to the visible spectrum to the total area of the curve as a

measure of the efficiency of the source as a light giver. Measured in this way the efficiency of an arc lamp is about ·1, that of an argand gas flame about ·016, and that of a carbon filament electric lamp about ·06.

126. Line Spectra.—In the preceding section we have been dealing with the spectra obtained when a solid body is heated so that it radiates, in which case the spectrum obtained is continuous. A similar result will be obtained with a molten solid, such as molten platinum.

When, however, the light given out by glowing gases or vapours is examined, the spectrum produced is of an entirely different character. Thus, if a salt of either of the metals sodium, calcium, strontium, lithium, &c., is held in a colourless flame, such as that of a Bunsen burner, and the light is examined in a spectroscope, the spectrum will be found to be no longer continuous, but to consist of a number of bright lines in various parts of the spectrum. The position and number of these lines varies for the different metals, but does not depend either on the salt of the metal used (chloride, bromide, sulphate, &c.) or on the nature of the flame into which the salt is introduced. The *number* of lines visible with any given metal depends, to a certain extent, on the temperature of the flame, but although new lines may make their appearance as the temperature is raised, the position of the lines already present does not vary. In the case of gases, the spectrum is obtained by passing the spark from an induction coil (§ 164) through the gas which is contained in a rarefied condition in a tube of the shape shown in Fig. 234. In addition to line spectra, under certain conditions of pressure and temperature, the spectra of some gases exhibit bands of light, which with a small dispersion are generally sharply defined on one side, but shade off gradually on the other. With a high dispersion, these bands are seen to be composed of numerous lines packed close together.

FIG. 234.

The line spectra are probably due to the fact that the molecules of gases and vapours are capable of vibrating in certain natural periods, and hence give out waves of the same frequency as these natural modes of vibration. Although at first sight the arrangement of the lines in the spectrum of a gas or vapour may appear quite irregular, yet a study of this subject has in recent years shown that in many cases certain relations exist between the frequencies of the vibrations corresponding to the different lines.

Since the positions of the chief lines in the spectra of the different elements are known, it is possible to identify the elements which produce the lines in the spectrum of an unknown substance. This spectroscopic method of analysis is in many cases a very delicate one, and further it

enables us to study the composition of bodies, such as the sun and stars, where the ordinary chemical analysis is not available.

127. Absorption and Emission.—When radiation falls on a body, part of the radiation is reflected at the surface, part (in the case at any rate of transparent bodies) is transmitted, and the rest is absorbed by the body. The proportions of the radiation reflected, transmitted, and absorbed vary greatly in different bodies. Thus a black body like lamp-black absorbs nearly the whole of the incident radiation, none being transmitted and very little reflected. Further, the proportions may vary greatly for the same body according to the wave-length of the radiation. Thus water which is very transparent to the visible radiation is very opaque to the ultra-violet, while ebonite is fairly transparent to the infra-red but quite opaque to the visible part of the spectrum. Hence water and ebonite are said to exhibit selective absorption. In the same way we have selective reflection, as when white light falls on a piece of red sealing-wax.

If the radiation given out by a body is received on a perfect absorber, that is a body which neither reflects or transmits any of the radiation, the temperature of the body will rise, and from this rise of temperature the amount of radiation received can be calculated. Although no known substance is a perfect absorber, yet lamp-black so nearly fulfils the conditions, that a metal surface coated with lamp-black is generally employed, a correction being applied to allow for the small proportion of the radiation which is reflected. If the amount of radiation given out under similar conditions by surfaces of different materials are measured in this way, the ratio of the radiation for any given surface to that of lamp-black (strictly that of a full radiator) is called the *coefficient of emission* of the surface. As will be seen from the following table, the coefficients of emission vary greatly.

Coefficients of Emission at 100° C.—

Lamp-black	1·00
Burnished silver	0·02
Gold-leaf	0·04

The ratio of the amount of radiation absorbed by a given surface to that absorbed by a similar surface of lamp-black, is called the *coefficient of absorption* of the surface for the given kind of radiation.

Experiment shows that for any given substance the coefficient of absorption and emission are equal, good absorbers being good radiators.

The coefficients of absorption and emission are equal.

In order to prove directly, by experiment, that the coefficients of absorption and emission for any given substance are equal, the apparatus shown in Fig. 235 has been devised by Ritchie. The two hollow metal drums B and C are

filled with air, and are connected by a glass tube D which is partly filled with some liquid, such as sulphuric acid. The drum A is also hollow, and can be filled with hot water. The faces of B and A, turned towards the right, are coated with lamp-black, and the faces of A and C, turned towards the left, are coated with silver-foil. The position of the liquid column having been noted when the whole instrument is at the same temperature, hot water is placed in A, and it is found that the liquid column does not move, showing that the drums B and C are receiving the same amount of heat from A. Now the drum B receives the heat emitted by a silver surface, the heat being absorbed by lamp-black, while the heat received by C is emitted by a lamp-black surface and absorbed by a silver surface. The heat received being the same, it shows that although the quantity of heat emitted by the silver surface is small, yet the lamp-black absorbing all this heat, the result is the same as when the large amount of heat radiated by the lamp-black surface falls on the silver surface, for in this case only a small proportion of the incident radiation is absorbed.

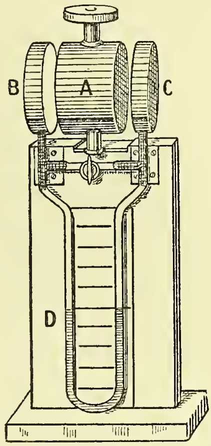


FIG. 235.

Not only are the emissive and absorptive powers the same for a given body when we consider the total radiation, but Kirchhoff first showed that the emissive and absorptive powers are the same for radiation of any given wave-length. Thus if a body shows selective absorption, *i.e.* absorbs light of a particular wave-length l to an abnormal extent, then when heated so that it gives out light it will give an excess of light of wave-length l. Thus if the light given out by a Bunsen flame in which a little common salt is placed is examined with a spectroscope, a very characteristic double bright line in the yellow will be observed. If now the light from the crater of an electric arc is examined with the spectroscope a continuous spectrum without any lines will be seen. On placing the flame tinged with salt between the arc and the slit of the spectroscope, the continuous spectrum will show a double *dark* line in exactly the place previously occupied by the bright ones produced by the flame alone.[1] The sodium flame has absorbed from the light of the arc, light of exactly the wave-length which it itself emits. Of course the flame itself will send a certain amount of light into the instrument, but compared to the bright adjacent parts of the spectrum of the arc, the two bands which only possess the light received from the flame look quite dark.

Reversal of spectral lines.

[1] Under these circumstances the lines are said to be *reversed.*

When sunlight is examined by a spectroscope, the spectrum is seen to be traversed by a large number of fine dark lines, which are called Fraunhofer's lines (see Fig. 215). These dark lines are due to absorption, chiefly in the sun's atmosphere, though a few are due to absorption in the earth's atmosphere. Since each dark line corresponds to what would be a bright line if the substance causing the absorption were itself emitting light, by comparing the positions of the Fraunhofer lines with the bright lines due to the different elements we can determine whether these elements are present in the sun's atmosphere.

Colours due to absorption.

The colours of most natural objects are due to absorption. Thus a piece of red glass looks red because it absorbs light of other wave-lengths. In the same way a blue pigment absorbs the whole of the red and yellow, but reflects the blue, together with some of the green and violet. A yellow pigment in the same way absorbs the violet and blue and reflects the yellow and some of the green and red. If we mix a blue and a yellow pigment the only colour which is unabsorbed by the *two* is the green, and hence such a mixture looks green. If, however, we allow yellow and blue light to fall on the same surface, the resultant mixture of *light* does not appear green, but almost white. In the same way a mixture of green and red pigments appears almost black, for the light which is reflected by one is absorbed by the other. Green and red lights, on the other hand, produce yellow. The question as to the colours perceived when different coloured lights are mixed is rather beyond the scope of this book. It may, however, be mentioned that by mixing a red light with a green and a blue in suitable proportions, *all* the colours of the spectrum can be reproduced. When we say we reproduce a colour, say yellow, by mixing a red light with a green, it must be understood that although the mixture looks to the normal eye exactly the same as a yellow of the spectrum, yet if we examine the mixture with a spectroscope we at once see that it differs physically from the spectrum yellow, in that one contains light of two wave-lengths and the other only light of a single wave-length. Further, a mixture of different coloured lights which appears to match a given light of a single wave-length to the majority of persons will appear quite incorrect to certain people and *vice versâ*. These people who have abnormal colour-senses are generally said to be colour-blind.

CHAPTER VIII

POLARISATION

128. Polarised Light.—If we take a slice of a crystal of tourmaline cut parallel to the crystallographic axis, and pass a ray of light through it, part of the light will pass through, and will, with most specimens of tourmaline, be coloured greenish owing to selective absorption within the crystal, otherwise to the eye the character of the light appears unaltered, and remains of the same intensity if the tourmaline plate is rotated. If the light which has passed through one tourmaline plate is allowed to fall on another, placed with its axis parallel to the first, the light will pass through the two; the only visible effect will be to *slightly* darken the greenish tint, the intensity being very slightly diminished by the second plate. If, however, the second plate is gradually rotated round an axis parallel to the light, so that the axes of the two crystals are inclined at a finite angle to one another, the intensity of the transmitted light will gradually diminish, till, when the axes are at right angles, none of the light which has passed through the first plate will pass through the second.

Hence the light which has passed through a plate of tourmaline has acquired properties which it did not before possess, in that it can no longer pass through a second plate of that substance when this plate is turned so that its axis is perpendicular to the axis of the first plate.

In order to see to what conclusions this experiment leads, let us consider an analogous problem. If we have a stretched string, we have seen that it is capable of two distinct modes of vibration, namely, a longitudinal vibration, in which the particles of the string move backwards and forwards in the direction of the length of the string, and a transverse vibration, in which the particles move in planes perpendicular to the length of the string. In the case of the string vibrating longitudinally, the appearance of the string is the same on all sides, *i.e.* it remains stretched straight between its extremities. When it is vibrating transversely, however, its appearance is ordinarily different on different sides, since it vibrates in a single plane. Hence a string vibrating transversely has definite sides, so that, to define its motion with reference to the surrounding medium, we must state the plane, passing through the undisturbed position of the string, in which the vibration takes place.

Another kind of transverse vibration of which a string is capable is that in which each particle describes a circle in a plane at right angles to the undisturbed position of the string. Suppose, then, that we cause a string to vibrate in this manner by attaching one end to a hook fixed at a little way from the centre of a rapidly rotating disc A (Fig. 236). The string will appear to swell out into something like the shape shown by the dotted lines. If now the string is passed through a narrow vertical slit D, since the motion of the string can then only take place up and down this slit, beyond D the motion will consist of transverse vibrations executed in a plane passing through the undisturbed position of the string and the slit, and by rotating the slit the plane in which the vibrations are taking place will also be rotated. Next, if a second slit is placed at F, and this slit is parallel to the first, the motion of the string, being parallel to this slit, will be unaffected. If, however, the first slit remaining vertical, the second slit F is turned out of the vertical, it will begin to interfere with

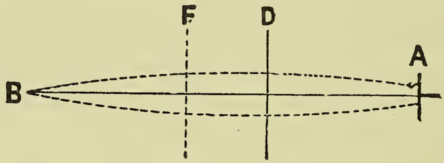


Fig. 236.

the vibration of the cord, and when it is horizontal it will no longer allow any of the motion of the cord, which is in a vertical plane, to pass, and hence the portion of the cord between the second slit and B will remain at rest.

The experiment with the crossed tourmalines gives just such a result as the above, and so we conclude that the reason the light will not pass through the second tourmaline, when the axes are at right angles, is that during its passage through the first the light vibrations have acquired sides, or, in other words, they now occur in one plane, so that they are stopped by the second tourmaline, just as the transverse vibrations of the cord are stopped by the second slit after they have been confined to one plane by the first.

Since no such action could take place with the cord vibrating longitudinally, we conclude that the light vibrations are transverse.

Ordinary light, then, consists of transverse vibrations, and since when a *single* tourmaline is used the intensity of the transmitted light does not change as the tourmaline turns, the vibrations must take place in all directions at right angles to the direction of the ray. After the passage of the light through the tourmaline, however, the transverse vibrations

all take place parallel to some definite direction, and the ray is said to be *plane polarised.* Thus when a ray of light IO (Fig. 237) passes through the tourmaline plate AB, which is cut so that the axis of the crystal is parallel to AB, the transmitted light is plane polarised, *i.e.* the vibrations take place in one plane. According to the electro-magnetic theory of light a ray of light consists of a periodic transverse wave of electric displacement, which is accompanied by a periodic transverse magnetic field in a direction at right angles to the electric displacement. Since one of these is always accompanied by the other, just as in the case of sound-waves in air the displacement of the air particles is accompanied by variations of pressure, we need only refer to the wave of electric displacement.

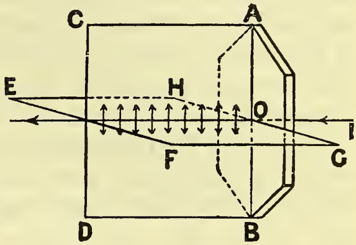


FIG. 237.

Plane of polarisation.

Thus a plane polarised ray of light is one in which the electric displacement is everywhere in the same direction, that is, everywhere perpendicular to a given plane EFGH (Fig. 237). This plane is called the *plane of polarisation* of the polarised light.

A ray of unpolarised light is such that the *direction* of the electric displacement is changed from time to time. Since the eye could not detect a periodic change of intensity which occurred about fifty times a second, and in this time about 10^{13} vibrations of light would occur, it is obvious that a variation of the direction of vibration which only occurred every few million vibrations could not be detected by looking at a source of light through a tourmaline.

129. Double Refraction.—When dealing with the refraction of light in Chapter II. we only considered isotropic media. If the medium through which the light passes has different properties in different directions, as is the case with many crystals, then the phenomena of refraction are more complex. As an example of a crystalline body we may take Iceland spar, which is a crystalline variety of calcium carbonate. It crystallises in various forms, but they all split most readily along certain planes which are always inclined to each other at fixed angles, so that by cleavage the crystals can always be reduced to the rhombohedral form shown in Fig. 238. The rhombohedron is bounded by six parallelograms, each of which has two acute angles of 78° 5′ and two obtuse angles of 101° 55′. Two of the solid angles, A and B (Fig. 238), are formed by three obtuse angles, the remaining six being formed

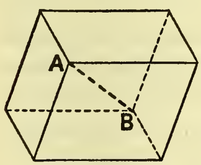


FIG. 238.

by one obtuse and two acute angles. A line drawn through either A or B, so as to be equally inclined to the three sides or edges which meet at the corner, is called the axis of the crystal. We shall find that the axis has very distinct and special optical properties, and since these properties are unaltered if the length, breadth, or thickness of the crystal are altered, as far as such optical properties are concerned, the axis must be looked upon as simply a direction in the crystal, so that all lines parallel to the crystallographic axis are optical axes.

If a ray of light is incident normally on one of the faces of a rhombohedron of Iceland spar, part of the light will pass straight through along POO′ (Fig. 239), just as would happen in the case of an isotropic body. Part of the light will, however, be refracted and travel along PEE′. Hence in this case there are *two refracted rays* corresponding to a single incident ray. This phenomena is spoken of as *double refraction*.

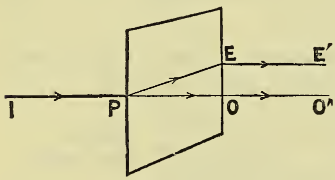


FIG. 239.

If the Iceland spar is turned round the line PO as an axis, and the transmitted light is received on a screen, the spot of light corresponding to the refracted ray POO′ will remain stationary, while that due to the ray PEE′ will rotate round the other as a centre, the line joining the two images being always parallel to the shorter diagonal of the parallelogram [1] which constitutes the face of the crystal on which the light is incident normally.

If the angle of incidence is not zero, then in general there will be two refracted rays, but while one of them obeys Snell's law in that the ratio of the sine of the angle of incidence to the sine of the angle of refraction is constant, whatever the direction of the incident ray, for the other ray this ratio is different for rays incident in different directions.

The refracted ray PO, which obeys the ordinary law of refraction, is called the *ordinary ray*, while the other ray PE, which does not obey this law, is called the *extraordinary ray*.

If the light which has been transmitted through the crystal of Iceland spar is allowed to fall on a plate of tourmaline, and if this plate is rotated, it will be found that in some positions only the ordinary ray is transmitted, and in others only the extraordinary ray. Since the positions for which the ordinary ray is extinguished are at right angles to the positions in which the extraordinary ray is extinguished, this experiment shows that both the

Ordinary and extraordinary rays.

[1] This is only true if the edges of the face are all of equal length. If this condition is not fulfilled, such an equilateral parallelogram must be marked out on the face.

ordinary and extraordinary rays are plane polarised, and that the planes of polarisation are at right angles to one another.

If in the case of a doubly refracting crystal a plane be drawn perpendicular to the face on which the light is incident, and so as to contain the optic axis, this plane is called the *principal plane* for the given face.

When the axis of the tourmaline plate is parallel to the principal plane of the spar, the ordinary ray is cut off by the tourmaline, which shows that the ordinary ray is polarised in a plane parallel to the axis of the tourmaline, *i.e.* parallel to the principal plane of the spar. When the axis of the tourmaline is at right angles to the principal plane the extraordinary ray is cut off, so that this ray is polarised in a plane perpendicular to the principal plane.

The properties of tourmaline are due to the fact that while an incident ray is split up into an ordinary and extraordinary ray as in the case of the Iceland spar, the tourmaline exerts a very powerful selective absorption on the ordinary ray, and but little absorption on the extraordinary ray. Thus a crystal of moderate thickness only allows the extraordinary ray to pass.

If a ray of light is incident on a crystal of Iceland spar in such a direction that the ray of light within the crystal is parallel to the optic axis, then the directions of the ordinary and extraordinary rays will coincide, that is, there will be only a single refracted ray, and for this direction in the crystal the refractive index is the same for both rays. The refractive index for other directions is less for the extraordinary ray than for the ordinary ray in Iceland spar. In some crystals, however, the reverse is the case.

Nicol's prism.

As a means of obtaining plane polarised light, a tourmaline plate is, for many purposes, unsuited, for, as has been mentioned, the light transmitted by tourmaline is coloured green. Since, when a beam of light is passed through a crystal of Iceland spar, two refracted beams are obtained, each of which is plane polarised, but in planes at right angles, if by any means we could intercept one of these refracted beams, the other would give us plane polarised light. Since the angular separation between the ordinary and extraordinary rays is not very great, it is not possible to stop one of the beams with a screen, unless only a very narrow beam is employed, or we use a very thick crystal.

The most convenient method of getting rid of one of the rays is to make use of total internal reflection for this purpose. A rhomb of Iceland spar is taken and cut in two by a plane AC (Fig. 240), perpendicular to the principal plane for the face AB. The two surfaces are then polished and cemented together in their original position by means of a thin film of Canada balsam.

Now the refractive index of Canada balsam (1·55) is greater than the minimum value for the extraordinary ray (1·486) in Iceland spar, and less than that for the ordinary ray (1·658). As total reflection can only occur when light is passing from a medium of greater to one of less refractive index, we can never get total reflection in the case of the extraordinary ray when passing from spar to balsam, so long as the ray passes in such a direction that the refractive index is less than 1·55. In the case of the ordinary ray, however, if the incidence is sufficiently oblique we shall obtain total reflection. Hence if the plane AC is suitably inclined, the ordinary ray, PO, will be incident on the surface AC at an angle greater than the critical angle, and will therefore be totally reflected along OO′, while the extraordinary ray, PEE′, will pass through the prism.

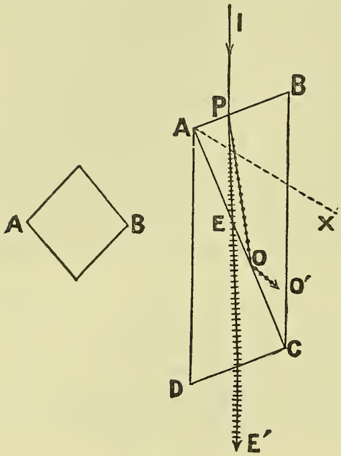


FIG. 240.

The light transmitted by such a rhomb of Iceland spar, which is called a Nicol's prism, will therefore be plane polarised, and since it is the extraordinary ray which is transmitted, the plane of polarisation is perpendicular to the principal plane, *i.e.* is a plane perpendicular to the paper in Fig. 240.

A Nicol's prism may be used, not only for producing plane polarised light, when it is called a polariser, but also for detecting whether light is plane polarised, and, if so, determining the plane in which it is polarised, when it is said to be used as an analyser.

If the light incident on the Nicol is unpolarised, then the intensity of the transmitted light will remain the same when the Nicol is rotated round the light ray as an axis, the intensity of the transmitted light being practically half that of the incident light. There is, however, a very slight loss due to reflection at E and P (Fig. 240), and where the ray leaves the crystal.

Double refraction produced by strain.

It is possible to render an isotropic body temporarily doubly refracting by subjecting it to a strain. This phenomenon can be examined by means of a bar of glass AB (Fig. 241) which is held in a metal frame, so that by screwing down the screw C the bar can be bent. If the bar is placed between crossed Nicols, so that the length of the bar is inclined at 45° to the principal planes of the Nicols, it will produce no effect so long as it is

unstrained. On bending the bar, however, the light which passes through the parts of the bar above and below the median line will be able to pass through the analysing Nicol, while the central line, shown dotted in the figure, remains dark as before.

When the bar is bent, the part above the dotted line is compressed and the part below is extended, while the central part is unstrained. Since the central part is unstrained, it produces no effect on the plane of polarisation of the light which passes through it, and this light is entirely cut off by the analyser. The strained parts, on the other hand, have become doubly refracting, and the incident plane polarised light is partly decomposed into light polarised parallel and at right angles to the length of the bar, that is, at 45° to the principal plane of the analyser, and so will be able to pass through.

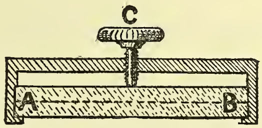


FIG. 241.

This method of placing a body between crossed Nicols, and seeing whether any light is then able to traverse the analyser, is a very delicate method of testing whether a transparent body is in a state of strain, and we shall see that under certain conditions even liquids may become doubly refracting due to strain.

130. Polarisation by Reflection. Rotation of the Plane of Polarisation.—If the light reflected from a non-metallic surface, such as glass, is examined with an analysing Nicol, it will be found that as the Nicol is rotated the intensity of the transmitted light varies. For a certain angle of incidence there is no light transmitted by the Nicol when its principal plane is parallel to the plane of incidence of the reflected light, while when the principal plane of the Nicol is perpendicular to the plane of incidence, the transmitted light is equal in intensity to the reflected light before it passes through the Nicol. This shows that, for this angle of incidence, the reflected ray is completely plane polarised in the plane of incidence. For all other angles of incidence the reflected ray is only partly polarised, *i.e.* consists of a mixture of ordinary unpolarised light with light which is polarised in the plane of incidence. The angle of incidence, for which the reflected beam is completely plane polarised, is called the polarising angle for the reflecting substance.

If, instead of consisting of ordinary light, the incident ray is plane polarised, and is incident at the polarising angle, then when the incident ray is polarised in the plane of incidence, *i.e.* the vibrations of the electric displacement are taking place perpendicular to the plane of incidence, the light will be reflected. If, however, the incident ray is polarised in a plane perpendicular to the plane of incidence, so that the vibrations

take place in this plane, none of the light will be reflected, but it will all be refracted into the reflecting substance.

If the incident light is polarised in intermediate planes, the reflected light will gradually increase in intensity as the plane of polarisation changes from the position in which it is perpendicular to the plane of incidence, to that in which it is parallel to the plane of incidence.

If n is the refractive index of a substance and θ is the polarising angle, then these quantities are related by the equation

$$n = \tan\theta \qquad . \quad . \quad . \quad . \quad (143)$$

This is known as Brewster's law.

If a ray of plane polarised light is passed through a solution of cane-sugar, and the light is then examined with a Nicol prism, it is found that as a result of passing through the solution the plane of polarisation has been rotated, the angle through which the plane of polarisation has turned depending on the strength and thickness of the solution. Substances which can in this way rotate the plane of polarisation are said to be *optically active.* Some substances rotate the plane of polarisation in one direction and some in the other. Even the same substance may rotate the plane in the two directions. Thus some samples of quartz rotate the plane in one direction, while other samples rotate it in the opposite direction, this difference being associated with a slight difference in the crystalline form. The amount of rotation produced per unit length is, however, the same for both varieties.

Rotation of the plane of polarisation.

BOOK V

MAGNETISM AND ELECTRICITY

CHAPTER I

MAGNETS AND MAGNETIC FIELDS

131. Magnetism. Magnets.—It was known to the ancients that certain ores of iron possess the property of attracting to themselves and retaining small particles of iron. This property was exhibited in a marked degree by some of the ores which came from a place in Asia Minor called Magnesia, and hence the ores which exhibited this property were called *magnetic stones*. All the phenomena connected with the properties of such magnetic stones, or magnets as they are now called, are referred to as magnetic phenomena, and the branch of physics dealing with this subject is called magnetism. The loadstone consists of equivalent proportions of the two oxides of iron, FeO, Fe_2O_3.

If a natural magnet, as a loadstone is often called, be dipped in iron filings, the filings will be found to adhere to the magnet in very characteristic tufts; these tufts are not uniformly distributed over the surface, but are much more marked at some parts of the surface, chiefly projecting corners, than at others.

In addition to the loadstone, there are other bodies which exhibit magnetic properties; chief among these are bars of hard steel which have been treated in a manner which we shall consider in detail in a subsequent section. Such a bar of steel is called an artificial magnet, but since we shall be dealing with artificial magnets exclusively we shall in future term it a magnet, and when it is dipped in iron filings it attracts them and forms tufts, but these tufts are almost exclusively confined to the two ends of the bar. The ends of the magnet where the power of attracting iron filings seems to be situated are called the poles of the magnet.

Another fundamental property of a magnet which also was known, at any rate to the Chinese, long ago is that when a magnet is suspended or pivoted so that it can turn freely about a vertical axis, it will set itself in a definite direction, which is very nearly parallel to the meridian, that is, it sets itself in the north and south direction. It is found that it is always the same end of any given magnet that points towards the north pole, and hence this pole of the magnet is called the *north-seeking pole*,

or simply the *north pole*,[1] while the other pole is called the *south-seeking pole*, or *south pole*. The fact that we are able in this way to distinguish the two poles of a magnet shows that there must be some difference between the two poles.

If a magnet is suspended by a fine thread, or pivoted on a point, so that it can turn freely in a horizontal plane, then, as we have already said, it will set itself in a direction which, in the absence of any disturbing force due to other magnetic causes, will be very nearly due north and south. If in these circumstances the north pole of another magnet is brought near the north pole of the suspended magnet, this latter will be repelled. If, however, the south pole of the magnet is brought near the north pole of the suspended magnet, this pole will be attracted. In this way we may verify the following general law: Two poles of similar name repel one another, while two poles of different name attract one another.

Magnetic field. Lines of force.

Since a magnet is capable of exerting a force on another magnet or of attracting a piece of soft iron even when at some distance, it follows that the space surrounding a magnet possesses some properties, due to the presence of the magnet, which it did not possess before the introduction of the magnet. Hence we speak of the space in which magnetic forces are exhibited as a *magnetic field of force*, or simply as a *magnetic field.*

If a small compass-needle is brought within the field of force of a magnet it will set itself at each point in a definite direction. If lines are drawn so that they are everywhere in the direction in which a compass-needle would set itself under the influence of the magnet, these curves are called the lines of force of the field of force of the magnet, or, more shortly, the lines of force of the magnet.

When a small pivoted compass-needle is placed in the neighbourhood of a magnet both its poles will be acted on by the two poles of the magnet. Thus the north pole of the needle will be attracted by the south pole of the magnet and repelled by the north pole, while the south pole of the needle will be attracted by the north pole of the magnet and repelled by the south pole. The needle will therefore set itself in such a direction that these four forces will have no resultant moment round the pivot about which the needle can turn. This will be the case when the needle sets itself with its length parallel to the resultant force acting on its poles. Hence the direction of the line of force through a given

[1] As we shall see, the reason a suspended magnet points in a north and south direction is that the earth behaves like a large magnet. Since the pole of this terrestrial magnet which is near the geographical north pole attracts the north-seeking pole of the suspended magnet, it must itself correspond to a south-seeking pole. Thus the magnetic pole of the earth which is near the geographical north pole is a *south* pole, and this has at times caused a little confusion.

point represents the *direction* of the resultant magnetic force at the point due to the two poles of the magnet which produces the field.

At any given point a north pole would be acted upon by a force tending to move it along the line of force in one direction, and a south pole would be acted upon by a force in the opposite direction. That direction along a line of force in which a north pole would tend to move is called the positive direction of the line of force. Hence the lines of force of a magnet start from its north pole and end at its south pole. A magnetic field in which the lines of force are all parallel is called a uniform magnetic field.

Any place at the surface of the earth is a magnetic field due to the earth's magnetism, the lines of force running from south to north. In the absence of magnets or magnetic bodies, the field due to the earth's magnetism is practically uniform over any moderate area.

If two magnets are so placed that their fields of force overlap, the resultant field will be such that its lines of force everywhere give the direction of the resultant magnetic force due to the two magnets.

132. Coulomb's Law. Unit Pole.—Although it is impossible to obtain either a north pole or a south pole alone, yet if we are dealing with a very long magnet in the space surrounding one end, the magnetic forces are almost entirely due to that pole alone, the other being so distant. Hence we can in this way get what amounts in practice to a single pole, and it will be found to save much circumlocution if we speak of a single pole and the forces which act on a single pole, as if it were possible to deal with such a single pole without the necessity of there also being a pole of the other kind.

By studying the forces between the poles of long thin magnets, Coulomb concluded that the force exerted between two poles is directly proportional to the product of the strengths of the poles, and inversely proportional to the square of the distance between them. This law enables us to define a unit in which to measure the strength of a pole. Thus a unit pole is such that two units placed at a distance from one another in air of one centimetre will exert a force on each other equal to a dyne. Thus if we have two poles of which the strengths are m and m' units, and they are at a distance d apart in air, Coulomb's law states that the force F one exerts on the other is given by

$$F=\frac{mm'}{d^2} \qquad . \quad . \quad . \quad . \quad (144)$$

If m and m' are of the same kind, *i.e.* both north poles or both south poles, the force is a repulsion. If they are of opposite kinds, the force is an attraction.

If we have a single north pole of strength m, then the lines of force will radiate out uniformly from the pole in every direction. If now we imagine a sphere of radius r described about the pole as centre, each of these lines of force will cut the sphere. If a unit north pole were placed at the surface of the sphere, the force acting upon it would by Coulomb's law be m/r^2. Now the force which would act on a unit pole when placed at a given point of a magnetic field may be taken as a measure of the strength of the field at that point. Hence at every point of the surface of the sphere considered we have that the strength of the magnetic field is m/r^2, and the direction of the field is normal to the surface.

As we have already seen, the lines of force indicate the direction of the field; they do not, however, show the strength of the field. It would obviously be an advantage if we could draw the lines of force in such a way that they indicate not only the direction, but also the strength of the field at every point. In order to secure this we may take the closeness with which the lines of force are drawn to represent the strength of the field. Thus in the case considered above let us draw through each unit of area of the sphere m/d^2 equally spaced lines of force, so that the strength of the field is numerically equal to the number of lines of force which pass through unit area taken at right angles to the directions of the lines of force at the given place. Since the area of the spherical surface is $4\pi d^2$, the total number of lines of force which cut through the surface is $4\pi m$. That is, with this convention the number of lines of force leaving a pole is 4π times the strength of the pole.

Strength of a magnetic field represented by number of lines of force per unit area.

The number of lines of force which cut through a given area is called the magnetic *flux* through that area, so that the flux per unit area taken *perpendicular* to the lines of force, or the flux density as it is called, is equal to the magnetic force at the surface considered.

A flux density of unity, that is, a field such that a unit pole experiences a force of one dyne, is called a *gauss*.

It is easy to see that in the case of an isolated pole of strength m where the lines of force are radial and uniformly spaced, that if we draw $4\pi m$ lines the number of lines per unit area taken at right angles to the lines of force will everywhere be equal to m/d^2, where d is the distance from the pole, and hence the relation between magnetic force and flux density given above must hold at all distances. It does not, however, follow that if in a given region of any magnetic field we draw the lines of force so spaced that the flux density is equal to the magnetic force in that region, that if these lines of force are simply prolonged their distribution will of necessity be such that *everywhere* the flux density is

equal to the magnetic force. Nevertheless, whatever the complexity of the field, this is the case, although a proof of this general proposition is beyond the scope of this book.

133. Couple acting on Magnet in a Magnetic Field. Magnetic Moment.—Suppose we had an ideal magnet which consisted of a north pole N (Fig. 242) of strength m, and a south pole S of strength m, at a distance l apart, the magnet being pivoted at the mid point O, so that it can rotate about O as an axis. If this magnet is placed in a uniform magnetic field of strength H, the direction of the field being Y′OY, so that the axis of the magnet makes an angle θ with the direction of the field, the magnet will tend to turn and set itself parallel to the lines of force of the field, and we now proceed to determine the couple tending to turn it.

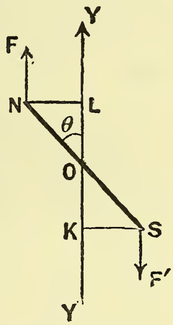


Fig. 242.

The pole N is acted upon by a force mH in the direction NF parallel to the lines of force of the field. The moment of this force about the pivot O is mH.NL or mH.ON sin θ.

In the same way the moment of the force acting on the pole S is mH.OS sin θ, tending to turn the magnet in the same direction. Thus the total turning moment, or the couple, acting on the magnet, is

$$Hm\{NO + OS\} \sin \theta$$

or

$$Hml \sin \theta.$$

If the magnet is placed with its axis at right angles to a field of unit strength ($\theta = 90°$ and $H = 1$) the couple will be ml.

Thus to calculate the couple we require to know both the pole strength m and the distance l between the poles. In any actual magnet it is found that the position of the poles is very indefinite, so that it is impossible in general either to determine m or l. If, however, a magnet is suspended in a known magnetic field and the couple tending to turn it is measured, we can use the relation given above to find the value of the product ml, while in general when dealing with problems involving magnets it is this product, rather than either m or l separately, which is required. Hence the product ml has received a special name, and is called the magnetic *moment* of the magnet. The magnetic moment may be defined as the couple which would act on the magnet when placed at right angles to a uniform field of unit strength.

Moment of a magnet.

Although we cannot actually measure either m or l, it is often convenient to speak of the pole strength of a magnet and the distance

between the poles when considering the action of one magnet on another. When, however, we come to use the expressions thus obtained we have to convert them into others in which the magnetic moments only are involved.

It will be noticed that it does not follow because two magnets have equal magnetic moments that their pole strengths and distance between the poles are equal. Thus a magnet of pole strength unity and distance between the poles 10 will have the same magnetic moment as another for which the pole strength is 10 and the distance between the poles is unity.

134. Field due to a Magnet. Couple due to the Action of one Magnet on another.—If we know the pole strength and distance between the poles of a magnet it is easy to calculate the direction and strength of the field at any point due to the magnet. Even if we do not know the exact position of the poles we may obtain the value of the field at points which are not too near the magnet with a considerable degree of accuracy.

We proceed to obtain expressions from the field in two simple cases. In the first place we require to find the strength of the field at points along the prolongation of the axis of a magnet. Let NS (Fig. 243) be the magnet of which the distance between the mid point P and either pole is L. If m is the strength of either pole, the field at a point O at a distance D from the mid point of the magnet is the resultant of

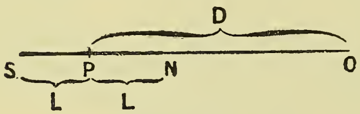


FIG. 243.

$$\frac{m}{(D-L)^2} \text{ due to the pole N and}$$

$$-\frac{m}{(D+L)^2} \text{ due to the pole S, the direction being}$$

the line PO. Thus the resultant field F at O is given by

$$F = m\left\{\frac{1}{(D-L)^2} - \frac{1}{(D+L)^2}\right\}$$

or
$$F = \frac{m.4L.D}{(D^2-L^2)^2}$$

But $2mL$ is the magnetic moment M of the magnet. Hence

$$F = \frac{2M.D}{(D^2-L^2)^2} \quad . \quad . \quad . \quad . \quad (145)$$

If D is very great compared to L, we may neglect the term L^2 in the denominator, so that

$$F = \frac{2M}{D^3} \quad . \quad . \quad . \quad . \quad . \quad (146)$$

Next let us calculate the value of the field at a point on a line drawn through the mid point of the magnet and perpendicular to the axis of the magnet. If a unit north pole were placed at O (Fig. 244), it would experience a force in the direction OA equal to $\frac{m}{(NO)^2}$ due to the pole N. It would also experience a force in the direction OS equal to $\frac{m}{(SO)^2}$ due to S. Since NO = SO, the above two forces are equal in magnitude. Hence if we take OB to equal OA and complete the parallelogram, the

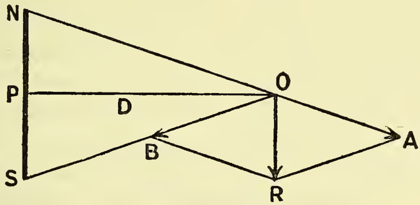


FIG. 244.

resultant force will be represented by OR. Since the triangles OAR and NOS are similar, we have

$$\frac{OR}{OA} = \frac{NS}{NO}$$

Hence $$OR = \frac{NS \, . \, OA}{NO}$$

But OA is equal to $m/(\text{NO})^2$, and hence OR, or F the field at O, is given by

$$F = \frac{mNS.}{(NO)^3}$$

or, since m.NS is the magnetic moment of the magnet,

$$F = \frac{M}{(NO)^3}$$

If as before the distance OP is called D and the distance PN or PS is called L, so that $(NO)^2 = D^2 + L^2$

$$F = \frac{M}{(D^2 + L^2)^{\frac{3}{2}}} \quad . \quad . \quad . \quad (147)$$

If D is very large compared to L this reduces to

$$F = \frac{M}{D^3} \quad . \quad . \quad . \quad . \quad (148)$$

By comparing equations (146) and (148) it will be seen that the field at a distance D from a magnet is twice as great on the prolongation of the axis as on a line drawn through the mid point perpendicular to the axis.

Couple exerted by one magnet on another.

Suppose a small pivoted magnetic needle of which the magnetic moment is M' is placed at O (Fig. 243) with its axis at right angles to PO, that is, with its axis at right angles to the field due to the magnet NS. Then, if this needle is very short, the field in which it finds itself will be practically uniform, and hence the couple which acts upon it is $M'F$, where F has the value given by (145) or (146) above.

In the same way, if the needle is placed at O (Fig. 244) with its axis along the line PO, it will experience a couple equal to $M'F$, where F has the value given by (147) or (148) above. The couple in this case will have half the value it would have in the former case. Now *experiment* has shown that the couple in one case is exactly twice as great as in the other. Hence we conclude that the hypothesis, that is, Coulomb's law, on which we have founded the above calculations is true. It can be shown that no other law than the inverse square of the distance law would give the above relation, so that we thus obtain an experimental verification of Coulomb's law, which does not necessitate our attempting to work with single poles, and is much easier to carry out with accuracy than Coulomb's original arrangement.

Deflection of pivoted needle by a magnet.

A pivoted or suspended magnetic needle may be used to compare the magnetic moment of a magnet with the strength of a magnetic field produced by some other cause, as, for instance, that of the earth.

Thus let *ns* (Fig. 245) be the pivoted needle of moment M' and NS a magnet of moment M, which is placed with its axis at right angles to the lines of force of the field, so that the centre of the needle is on the prolongation of the axis. Let the strength of the field be H and its direction parallel to OA. The needle will then be acted upon by a couple due to the magnet NS and another couple due to the field H, and it will take up a position inclined at an angle θ to the lines of force of the field such that these couples are in equilibrium. The couple

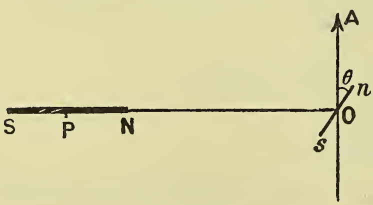


FIG. 245.

tending to turn the needle in the anti-clockwise direction due to the field is $M'H \sin \theta$, while the couple due to the magnet tending to turn the needle in the opposite direction is $\frac{M'M}{2D^3} \cos \theta$, for the angle between the field due to the magnet and the axis of the needle is $90° - \theta$ and $\sin(90° - \theta) = \cos \theta$. Hence, since the needle is in equilibrium,

$$M'H \sin \theta = \frac{M'M}{2D^3} \cos \theta$$

or

$$\frac{M}{H} = \frac{D^3}{2} \tan \theta \quad . \quad . \quad . \quad . \quad (149)$$

It will be observed that θ is independent of the magnetic moment of the needle, and that by measuring θ and the distance D we can then determine the ratio of the magnetic moment of the magnet NS to the strength of the field before this magnet was introduced. We cannot, however, from this experiment deduce the value of either M or H. To be able to do this we require to find some other relation between these two quantities, and in the next section such a relation is shown to be obtained if the period of the magnet NS when vibrating in the field H is measured.

A similar expression to (149) can be obtained when the needle occupies a position such as O (Fig. 244) with reference to the deflecting magnet.

135. Time of Vibration of a Magnet when suspended in a Magnetic Field.—If a magnet of which the moment is M, is suspended by a fine fibre in a horizontal magnetic field of strength H, it will set itself parallel to the lines of force of the field. If the magnet is then deflected so that its axis makes an angle θ with the field, it will be acted upon by a couple $MH \sin \theta$ tending to *restore* it to its original position, and if released it will execute vibrations about this position.

If the angle θ through which the magnet is deflected is small, we may take $\sin \theta = \theta$, and hence the restoring couple is $MH\theta$. Hence if the moment of inertia of the magnet about the suspending fibre is I, we have from equation (94) that the period of the vibrations is given by

$$T = 2\pi\sqrt{\frac{I}{MH}} \quad . \quad . \quad . \quad . \quad (150)$$

Since in this case, as in that of the simple pendulum, the restoring couple is proportional to the *sine* of the angle of deflection, the period is only independent of the amplitude of the vibrations when this amplitude is small.

CHAPTER II

TERRESTRIAL MAGNETISM

136. The Earth's Magnetic Field.—We have already mentioned that at all points on the earth's surface there exists a magnetic field. This field varies in magnitude and direction from place to place, and to completely specify the value of the field at any given point we require to know (1) the direction of the lines of force of the field, and (2) the strength of the field. As will be shown presently, the *direction* of the earth's field is not in general horizontal, the lines of force in the northern hemisphere dipping downwards, hence it is often convenient to suppose the earth's field resolved into two components, one of which is horizontal and the other vertical. Since the strength of a magnetic field is the *force* which would act on a unit pole, and we may resolve this force into components as described in § 10, a magnetic field may be resolved into component fields, and the relation between the components and the original field will be the same as those between two component forces and their resultant, that is, we compound and resolve magnetic fields by the parallelogram law. The horizontal component of the earth's field is sometimes called the *horizontal force*, sometimes the *horizontal intensity*, and sometimes simply the *horizontal component*, and is generally indicated by the letter H. Similarly the vertical component is known as the *vertical force* or the *vertical intensity*, and is indicated by the letter V. The actual field, of which the above are the components, is called the *total force* or *total intensity*.

The angle which the direction of the actual field (total intensity) makes with the horizontal is called the *dip*.

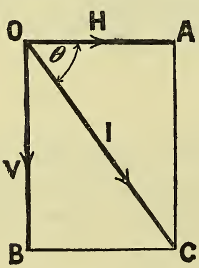


FIG. 246.

The relation between H, V, and the total intensity I at any place can at once be obtained from Fig. 246. If OC represents the total intensity in magnitude and direction, and OA and OB represent the horizontal and vertical components, the angle AOC is the dip, θ. Hence we have the following relations—

$$\frac{H}{I} = \cos\theta\,;\ \frac{V}{I} = \sin\theta\,;\ \frac{V}{H} = \tan\theta \quad . \quad (151)$$

To completely define the horizontal component we require not only

to know its magnitude, but also the *direction* in which it acts, that is, we require to know the angle it makes with some fixed direction taken in a horizontal plane. The fixed direction universally employed is the geographical meridian, and the angle between the direction of the horizontal component and the geographical meridian is called the *declination* or *variation.*

The horizontal component, total intensity, and the vertical component must of necessity lie in the same plane, and this plane is vertical and is known as the magnetic meridian. The angle between the magnetic and geographical meridians is the declination.

The declination is of great practical importance since a compass-needle gives the direction of the magnetic meridian, and hence to deduce the true geographical north from the indications of a compass we must know the declination. The declination varies from point to point on the surface of the earth, and maps have been constructed on which the declination at the different points are shown by means of lines which pass through all places at which the declination has the same value. Such lines are called *isogonal* lines. The particular isogonal which corresponds to points where the declination is zero, *i.e.* points where the compass points due north, is called the *agonic* line. At the present date one agonic line runs down the east side of North America, cuts off a part of South America, and then passes to the Antarctic Ocean. Reappearing on the other side of the south pole, it passes through the extreme west of Australia, through the Persian Gulf, and finally enters the Arctic Ocean near the North Cape. There is a second agonic line which forms an oval, the greater part of which lies in Siberia.

Lines drawn through the places where the dip has a given value are called *isoclinals.* The line of zero dip is called the magnetic equator, and forms a circle which follows approximately the geographical equator. In the southern hemisphere the south pole of a freely suspended magnet dips.

At two points of the earth's surface the dip is 90°, *i.e.* the lines of force of the earth's field are vertical. These points are called the magnetic poles, and one of the agonic lines passes through these poles. Since when the dip is 90° the horizontal component is zero, at the magnetic poles the compass-needle will not point in any fixed direction, but will remain in any azimuth in which it may be set.

Not only does the earth's field change in magnitude and direction from place to place, but at any given place the value is found to alter from year to year. This gradual change is called the secular change or secular variation. The rate of the secular variation is very different at different places, and in the case of London the declination has changed

from 11° east in 1570 to 24° west in 1810, and is now (1912) 15° 50′ west, and decreasing at the rate of about 6 minutes per annum.

In addition to the slow secular change, a regular periodic change of small amplitude takes place daily, called the diurnal variation, and annually, called the annual variation. Occasionally disturbances take place in which the earth's field undergoes sudden changes, which may go on for several days, the field, however, in time returning to its normal value. These are called magnetic storms, and are particularly liable to occur when the aurora borealis is visible.

The cause of the earth's magnetic field and of the variations which occur is still unknown.

137. Determination of the Magnetic Elements.—Since if we know the horizontal component, the declination, and the dip, we can by means of the relations given in equation (151) calculate the vertical component and the total intensity, the value of the earth's field at a given spot is fully determined if we measure (1) the declination, (2) the dip, and (3) the horizontal component, these three quantities being known as the magnetic elements at the place.

The declination being the angle between the direction of the horizontal component and the geographical meridian, the measurement involves two observations—(*a*) the magnetic meridian, and (*b*) the geographical meridian. The magnetic meridian is obtained by observing the direction of the axis of a magnet which is suspended by a fine thread so that it can turn freely about a vertical axis. The practical difficulty of this part of the measurment is caused by the fact that the magnetic axis of a magnet does not generally exactly coincide with the geometrical axis.

Determination of the declination.

The magnet usually employed consists of a hollow steel cylinder A (Fig. 247), which is fixed in a brass collar to which are attached two pegs, B and C, either of which fits into a clip attached to the end of a fine thread formed of unspun silk. At one end of the hollow magnet is placed a fine scale, S, engraved on a piece of glass, while at the other end is placed a lens L. The focal length of this lens is equal to the length of the cylinder, so that the rays of light proceeding from any point in the scale S leave the lens as a parallel pencil. The line joining the central division of the scale and the optical centre of the lens is taken as the geometrical axis of the magnet.

If AB (I., Fig. 248) is the plan of a magnetic needle suspended by a fine thread attached at C, and of which the magnetic axis is *ns*, then it will set itself with the magnetic axis in the magnetic meridian NS. In this case the geometrical axis of the needle points to the west of the magnetic meridian. If the needle is reversed, so that what was the lower

side is now the upper, then, as is shown at II., the geometrical axis AB will point as far to the east of the magnetic meridian as it did before to the west. Hence the magnetic meridian is half-way between the positions of the geometrical axis before and after the reversal of the magnet.

The cylindrical magnet shown in Fig. 247 is suspended in a box fixed to the centre of a divided circle, while an arm attached to the circle

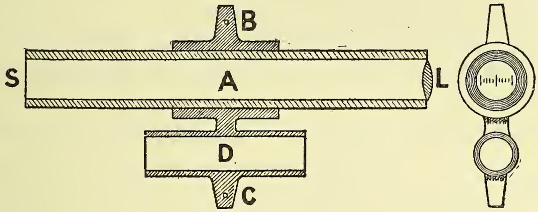


Fig. 247.

carries a small telescope, in the eye-piece of which are two intersecting cross-wires. The telescope is turned till the centre division of the scale, S, coincides with the vertical cross-wire, first when the magnet is suspended by B, and then when it is reversed and is suspended by C. The mean of the readings on the divided circle corresponding to these two positions gives the reading corresponding to the magnetic axis of the

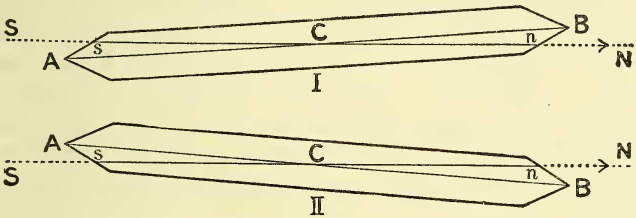


Fig. 248.

magnet. The reading corresponding to the geographical meridian is obtained either by turning the telescope to view some object the bearing of which is known, or by observing the time of transit of a star or the sun over the vertical cross-wire.

When determining the declination by means of a magnet suspended by a thread, the effects of gravity on the magnet do not influence the observations, for the weight of the magnet will have no effect in producing a rotation about the thread as an axis.

In order to determine the dip, however, we have to support a magnet

so that it can turn freely about a *horizontal* axis, and then measure the angle which its magnetic axis makes with the horizontal. If the axis about which the magnet is allowed to turn passes through the centre of gravity of the magnet, the weight will have no moment round this axis, and will therefore not affect the position of the magnet. Since, however, it is practically impossible to secure this condition, the observations have to be so arranged that errors due to small departures from this condition may be eliminated. The principle is similar to that employed in the case of the determination of the declination, viz. to take readings in pairs, such that the error in the separate readings affects the result in the opposite way, and hence the mean of the two readings gives the true value.

Determination of the dip.

Suppose AEDB (Fig. 249) is the needle, the axle being at C, while the centre of gravity is at G, a point which does not coincide with C. We may consider the effect of this displacement of the centre of gravity as split up into two parts, one a displacement along the axis of the needle to K, and the other a displacement at right angles to the axis to H.

First consider the displacement of the centre of gravity at right angles to the axis of the needle, *i.e.* to H. If the end A is dipping, that is, in the northern hemisphere, if A is a north pole, and the needle is placed in its bearings, so that H is above C, the effect of the weight of the needle will be to increase the measured dip, while if the needle is reversed in its bearings, so that H is below C, the weight will decrease the measured dip by the same amount. Hence the mean of the readings obtained will be free from error due to the displacement of the centre of gravity at right angles to the axis.

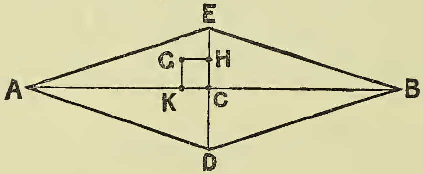


FIG. 249.

As long as the end A is dipping, the displacement of the centre of gravity along the axis to K will always increase the measured angle of dip. If, however, we remagnetise the needle, so that the end B dips, then the displacement of the centre of gravity to K will decrease the measured dip, so that by reversing the polarity of the needle this error can be eliminated. The fact that the magnetic axis of the needle may not coincide with the line joining AB is eliminated by the reversal of the needle when eliminating the effect of the displacement of the centre of gravity perpendicular to the axis, for the same reasons as in the case of the declination.

In order to measure the angle of dip, the needle is placed with its

axle resting on two small horizontal agate knife-edges, so that the axle is at the centre of a graduated circle, any slight want of agreement between the position of the axle and the centre of the circle being eliminated by reading the position of both ends of the needle by means of the two microscopes.

The only remaining source of error which may occur, owing to the imperfect adjustment of the instrument, is that due to the fact that the line joining the zero gradations on the circle may not be truly horizontal. The error due to this cause can be eliminated by taking two sets of readings with the instrument turned so that the graduated side of the circle faces first east and then west. For the measured dip will in one case be greater than the true dip, and in the other case less by the angle which the line joining the zeros makes with the horizontal.

The angle between the magnetic axis of the needle and the horizontal is only equal to the dip when the vertical plane in which the needle can turn coincides with the magnetic meridian, that is, when the axle of the needle points due magnetic east and west. In order to secure this condition, the circle is turned until the needle is vertical. The needle being vertical shows that in the plane in which it can move there is no horizontal component of the earth's magnetic field. Now obviously this can only occur when the plane in which the needle moves is at right angles to the direction of the horizontal component, that is, at right angles to the magnetic meridian. We may also see that this must be so from the following reasoning: When the axle about which the needle turns is parallel to the horizontal component, this component can have no moment tending to produce rotation about the axle. Hence, when the plane in which the needle turns is at right angles to the magnetic meridian, the vertical component is the only one which has a directive influence on the needle, which therefore sets itself in a vertical position. The position of the circle when the needle is vertical is read off on a horizontal divided circle attached to the stand, and then by means of this same horizontal circle the vertical circle, and with it the uprights carrying the needle, is turned through 90°, which brings its plane into the magnetic meridian.

Measurement of the horizontal component.

The horizontal component is measured by determining the period of oscillation of the magnet used in measuring the declination, which gives the value of MH where M is the magnetic moment of the magnet. This magnet is then used to deflect another suspended magnet, and from the deflection of this needle the value of M/H is obtained (see p. 389). By means of these two relations the value of H can then be obtained.

In a certain number of observatories a continuous record of the values

of the magnetic elements is kept by means of self-recording instruments.

Continuous record of changes in magnetic elements.

The records are obtained by means of the trace left by a spot of light reflected from a mirror attached to a magnet on a sheet of photographic paper, which is kept in uniform movement by means of clockwork. In the case of the declination, the mirror is simply attached to a magnet which is suspended by a long fine thread, so that it can turn freely about a vertical axis, and so, by always setting itself in the magnetic meridian, gives a record of the changes that take place in the position of this meridian, that is, shows the changes in the declination.

The changes in the horizontal force are recorded by means of a magnet which is suspended by a quartz fibre. The top of the fibre is turned till the magnet sets itself at right angles to the magnetic meridian, under which conditions the earth's field exerts a turning couple on the magnet equal to MH (§ 133), this couple being balanced by the couple due to the fibre. If the value of the horizontal force H alters, the couple due to the magnetic forces alters also in the same proportion, and the magnet turns about a vertical axis till the couple due to the fibre becomes equal to the new couple due to the magnetic forces. Changes in the declination will, however, not affect the position of the magnet, since it is at right angles to the magnetic meridian.

Since no satisfactory way of recording the changes that take place in the dip has been devised, it is usual to record the changes in the vertical force. For this purpose a magnet is balanced on knife-edges in such a way that it is in an approximately horizontal position. If, say, the vertical force decreases, then the downward force acting on the north pole and the upward force acting on the south pole both decrease, and hence the north pole of the balanced magnet rises and the south pole falls, just as when, in a balance, the load of one pan is increased and that of the other is decreased. The motions of such a balanced magnet will therefore indicate the changes that take place in the vertical force, and since the magnet with the fibre suspension gives the changes that take place in the horizontal force, the changes in dip and in the total force can be immediately calculated from the records given by the two instruments.

CHAPTER III

MAGNETIC INDUCTION

138. Induced Magnetism in Iron.—If a piece of soft iron is placed in a magnetic field it becomes a magnet, and the lines of force of the field crowd into the iron. The part of the iron where the lines of force enter is a south pole, while the part where the lines of force leave the iron is a north pole. On removal of the iron from the magnetic field, most of the magnetism of the iron disappears. If a bar of hard steel is treated in the same manner it will become magnetised, but most of the magnetism will remain after the removal of the field. Other substances besides iron become slightly magnetised when placed in a magnetic field, but with the exception of an alloy of manganese, aluminium, and copper, called Heusler's alloy, together with nickel and cobalt, the effect is very slight compared to that in the case of iron. A bar of iron, if suspended so as to be able to turn freely in a magnetic field, will turn in such a way that its longest dimension is parallel to the lines of force. Iron, nickel, cobalt, &c., and the other magnetic materials which behave in a similar manner to iron, are called *paramagnetic*. Such substances as bismuth and antimony, on the other hand, tend to turn so that the longer dimension is perpendicular to the lines of force of a magnetic field, and are called *diamagnetic*. Paramagnetic bodies are attracted by a magnet, while diamagnetic bodies are repelled. The effect in the case of diamagnetic bodies is, however, very feeble compared to that of such paramagnetic bodies as iron, nickel, and cobalt.

For all practical purposes the only substances of which the magnetic properties need be studied are the different kinds of iron, and we now proceed to consider the behaviour of iron when magnetised.

We have already shown how the magnetic field of a magnet can be represented by lines of force which leave the magnet near the north pole and enter the magnet near the south pole. Although these lines of force have no physical existence, yet they form a very convenient means of picturing to ourselves the conditions of a magnetic field. Further, it can be shown that if we imagine that the lines of force behave as if they were stretched elastic threads, which mutually repel one another, then all the attractions and repulsions between magnets and magnetised bodies would be accounted for. Thus we shall in the following pages speak as if the

lines of force were actual physical entities, and that a tension exists along the lines of force and a pressure at right angles, the number of lines of force being taken so that the number of lines at a given place which thread through unit area taken at right angles to their direction is equal to the magnetic force existing at that place (see p. 384).

We have hitherto spoken of the lines of force as if they started at the north pole of a magnet and ended at a south pole. Now, as we shall see later, magnetic lines of force exist in the neighbourhood of a wire in which an electric current is flowing, and these lines of force consist of closed curves which encircle the wire, the lines being continuous, having neither beginning nor end. The question then arises, are the lines of force of a magnet also continuous, *i.e.* should we look upon them as passing through the substance of the magnet from the south pole to the north pole. The probability that this is the case can be seen from the following considerations. Suppose we take a long thin steel rod which is magnetised uniformly, with a north pole at one end and a south pole at the other. The lines of force in the air surrounding the magnet will run from the north pole to the south pole. Now suppose we bend the magnet round so that the north and south poles are brought into contact. It is then found that the magnet has no field, *i.e.* there are no lines of force now passing through the air in the neighbourhood of the magnet. When, however, the poles are separated, a magnetic field is at once produced between the poles, *i.e.* lines of force now pass through the air. As it is unlikely the lines of force would only come into existence when the poles are separated, we infer that when the poles are in contact the lines of force exist, but that they simply pass round the ring of steel. Thus when the poles are separated we look upon the lines of force as continuous lines, which in the air run from the north pole to the south pole, and in the iron run from the south pole to the north.

Suppose we have a long thin cylindrical magnet of cross-section s, length l, and pole strength m. Then, as shown on p. 384, the number of lines of force which leave the north pole, or enter the south pole, is $4\pi m$.

Intensity of magnetisation.

Hence $4\pi m$ lines of force pass through the magnet, and the number of lines of force which thread through unit area of the section of the magnet is $4\pi m/s$. Thus the quotient m/s is proportional to the closeness with which the lines of force are packed in their passage through the magnet, and it is taken as a measure of the degree of magnetisation of the steel of the magnet, receiving the name of the *intensity of magnetisation.*

If I is the intensity of magnetisation we have $I=m/s=ml/ls$. But ml is the magnetic moment M of the magnet, and ls is the volume V. Hence

$$I=M/V \quad . \quad . \quad . \quad . \quad (152)$$

So that the intensity of magnetisation of a magnet may be defined as the magnetic movement per unit volume.

In the above we have been considering the magnetic state of the iron in the case of a *permanent magnet*. We now have to proceed to consider the case where the magnetism of the iron is induced owing to the fact that the iron has been introduced into a magnetic field caused by some external agency. Now we have reason to believe that each of the molecules of a piece of *unmagnetised* iron is really a little magnet, but that the axes of these little magnets are oriented in all directions. Each molecular magnet will have its lines of force, just like a large magnet, but owing to the juxtaposition of large numbers of these molecular magnets all turned in different directions, if we consider unit area within the iron as many lines will thread through the area in one

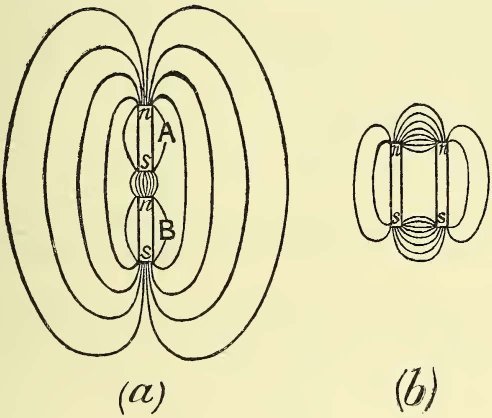


FIG. 250.

direction as in the opposite. When, however, the iron is placed in a magnetic field, the molecular magnets will be more or less completely turned so that their axes all turn in the same direction, and hence more of the lines of force of these magnets will now pass through a given area in one direction than in the opposite direction. For if two small magnets A and B (Fig. 250) are placed with their axes parallel and their north poles turned in the same direction, the lines of force will run somewhat as shown at (*a*) in the figure, and will extend to some distance from the pair. If, however, the magnets are placed as at (*b*) Fig. 250, the lines of force do not extend nearly so far into the surrounding space. It will thus be understood that when the molecular magnets are orientated by the action of the external magnetising field the distribution of the lines is altered, so that within the iron they tend to run in the same direction as the lines of the field.

Hence when a piece of iron is placed within a magnetic field we have to do with two sets of lines of force—(1) those which are due to the magnetising field and which existed before the iron is introduced, and (2) those due to the magnetism of the molecules of the iron itself.

Suppose that a long unmagnetised cylindrical bar of soft iron of cross-section s is placed in a uniform magnetic field of strength H, with its length parallel to the lines of force of the field. Then if the cylinder were of an unmagnetisable material, the number of lines of force due to the field which would cross the cross-section of the cylinder is sH. Owing, however, to the fact that the cylinder becomes magnetised by induction, there will be in addition a certain number of lines of force both within the cylinder and in the air outside, due to this induced magnetism.

Owing to the induction, poles will be induced at the ends of the iron, and these poles will in general produce a force within the material of the iron which will tend to diminish the strength of the inducing field. This disturbing action of the induced poles causes a considerable complication in the consideration of the problem, and so we shall at first consider the case of a very long cylinder of comparatively small cross-section. In this case, if we confine ourselves to a consideration of the state of the iron near the middle of the cylinder, the influence of the poles, which are by supposition at a considerable distance, may be neglected.

Magnetic Induction.

If m is the strength of the poles induced in the iron there will be $4\pi m$ lines of force due to this induced magnetism. Since that end of the cylinder which points in the direction in which the lines of force of the field run becomes a north pole, the induced lines of force will run in the air in the opposite direction to the lines of force of the field, but within the iron they will run in the same direction as the lines of force of the field. The number of lines of force which pass through the iron is therefore sH lines, due to the inducing field, and $4\pi m$ lines, due to the induced magnetisation, or $sH+4\pi m$ in all. Hence the number of lines of force which cross unit area of the cross-section of the iron is $H+4\pi m/s$. But the intensity of magnetisation of the iron is equal to m/s, for the lines of force due to the field alone have nothing to do with the magnetism of the material, and in fact remain the same whatever the nature of the material of which the cylinder is composed. Hence if I is the intensity of the magnetism induced in the iron, the number of lines of force which cross *unit area* of the cross-section of the cylinder is $H+4\pi I$. This quantity is called the *induction*, B, in the iron, so that

$$B=H+4\pi I \quad . \quad . \quad . \quad . \quad (153)$$

The total number of lines which thread through the iron, i.e. Bs, is called the magnetic *flux* through the iron, so that B is the flux density.

If the length of the iron bar is not very great compared to its cross-section the poles produced at the end of the bar due to the magnetisation of the iron will produce an appreciable effect, and the magnetising field within the bar will consist of two parts, one due to the external field and the other due to the induced poles. If it were possible to keep the induced poles in the place which they occupy and yet remove the iron, then the field due to the poles would be opposite in direction to that due to the inducing field, for the induced north pole is turned towards the direction in which the lines of force of the field run and the lines of force due to the two induced poles run from the north pole to the south. The field due to the induced poles therefore opposes the external magnetising field, and hence is called the *demagnetising field*. The demagnetising field can be approximately calculated in the case of fairly long cylinders, but in general it is better to use specimens which are either in the form of such long rods that the demagnetising field is negligible or in the form of rings in which the iron is magnetised parallel to the axis of the ring so that there are no poles developed.

The resultant of the field due to the external magnetising agent and that due to the induced poles is generally called the *magnetising force* for the specimen of iron being tested.

Magnetic induction B and magnetising force H are both of the same nature as field strength, and hence are measured in gausses.

Susceptibility and permeability.

The ratio of the intensity of magnetisation I to the magnetising force H by which it is produced is called the magnetic *susceptibility* of the iron, and is generally indicated by the letter k.

The ratio of the induction, B, to the magnetising force is called the *permeability* of the iron, and is generally indicated by the Greek letter μ.

Thus $$\frac{I}{H}=k \text{ and } \frac{B}{H}=\mu \quad . \quad . \quad . \quad (154)$$

Since $$B=H+4\pi I$$

we have $$\mu=1+4\pi k \quad . \quad . \quad . \quad . \quad (155)$$

The permeability of iron is a most important quantity in the design of electrical machinery, for the greater the permeability the greater the induction produced in the iron by a given magnetising force.

The relation between the induction B and the magnetising force H

for a sample of iron is given in Fig. 251. It will be noticed that with increasing H the value of B increases at first comparatively slowly, the curve being at first nearly straight. At A, however, the curve tends upwards, so that B increases very rapidly with increasing H. At B the increase becomes much

The B-H curve for iron.

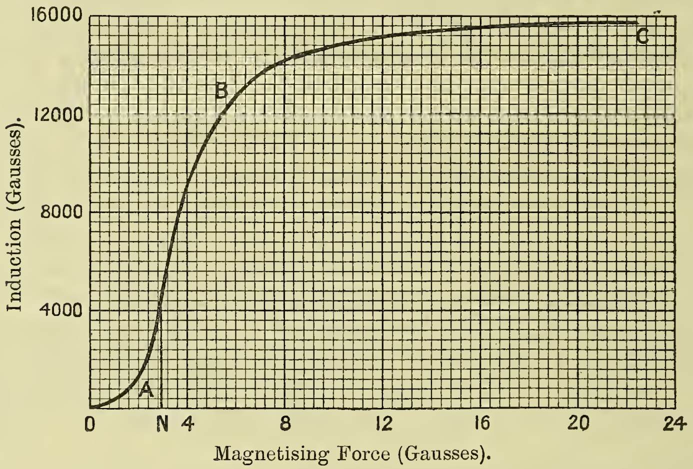


Fig. 251.

slower, till finally the curve becomes practically a straight line which is almost parallel to the x axis with the scales for B and H adopted in the diagram. If, however, as in Fig. 252, the scale adopted for H is the same as that for B, in place of being a thousand times greater as is the case in Fig. 251, the upper part of the curve is practically a straight line inclined at an angle of 45° to the axes, indicating that for large values of H the change in B is equal to the change in H. This of course indicates that the increase in the number of lines of force threading through the iron is *solely* due to the increase in the magnetising field, no *additional* lines being produced by the orientation of the magnetised iron molecules. When the iron reaches this condition the iron is said to be *saturated*, so that the

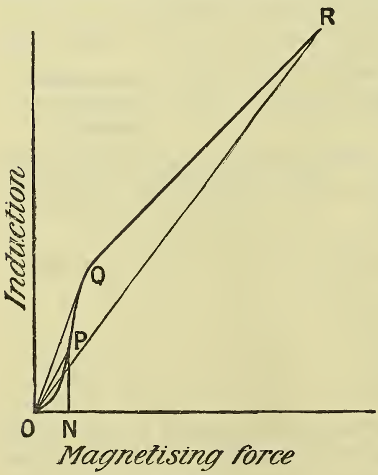


Fig. 252.

horizontal part, BC, of the curve in Fig. 251 corresponds to the saturated state.

If we take any point P (Fig. 252), on the B-H curve, the permeability which is B/H is represented by the ratio of PN to NO, that is, the permeability is equal to the tangent of the angle PON.

If from O we draw a tangent OQ to the B-H curve, the point Q corresponds to a value of the magnetising force for which the permeability is a maximum. For higher values of H the permeability gradually decreases, and the angle made by the straight line joining a point on the curve to the origin approaches nearer and nearer to the value 45°. That is, since $\tan 45° = 1$, the permeability approaches the value unity when saturation is reached.

The value of H required to produce saturation varies greatly according to the quality of the iron. Thus in the case of a very soft iron

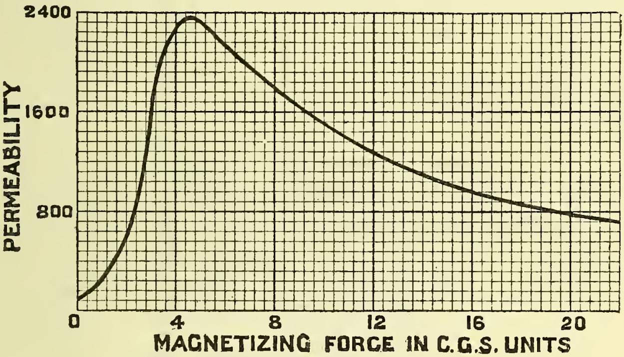


FIG. 253.

saturation may be reached when $H = 2000$ gausses, while with a hard steel 15,000 gausses may not be sufficient to produce saturation.

The relation between the permeability and magnetising force from the sample of iron from which Fig. 251 gives the B-H curve is given in Fig. 253. For very weak fields the permeability is almost constant. This corresponds to the first part of the B-H curve, which is almost a straight line.

When a piece of iron or steel is heated to a bright red it loses its power of becoming magnetised, or, if permanently magnetised, all this permanent magnetism will be lost. A similar change takes place in the case of nickel and cobalt.

Effect of temperature on the magnetic properties of iron.

The temperature at which a particular sample of a magnetic metal loses its magnetic properties is called the critical temperature for that metal.

With small values of the magnetising force the loss of the magnetic properties of soft iron, as the temperature reaches the critical

point, is much more sudden than with strong magnetising forces. In Fig. 254 the relation between the permeability of soft iron and the temperature, as obtained by Morris, for different magnetising forces is shown.

For low magnetising forces it will be seen that the permeability increases slowly with rise of temperature up to a temperature of about 600°; the increase then becomes very much more rapid, till at about 750° the curve becomes almost vertical. The permeability reaches a maximum value for a temperature of 775°. Above this temperature there is a sudden decrease of the permeability, and at a temperature of 785° the permeability is practically unity, that is, the iron has lost its magnetic properties. For large magnetising forces (of 4·0 gausses and

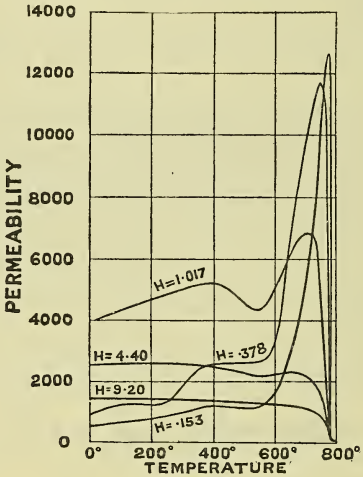


FIG. 254.

over) there is no increase of the permeability as the temperature increases, and the decrease of the permeability as the critical temperature is approached is very much more gradual and commences at a temperature of about 670°.

Similar results are obtained in the case of steel, although the loss of magnetic properties is not so sudden as in the case of wrought iron.

An examination of the cooling curve (§ 74) for iron shows a marked evolution of heat at the critical temperature, indicating that some molecular change takes place at this temperature.

We have above spoken of the induction in a sample of iron as if when the magnetising field has a given value the value of the induction were determinate. This, however, is only the case when we start with the iron in the unmagnetised condition and increase the magnetising force steadily to the given value. In general the value of B corresponding to

any given value of H depends to a considerable extent on the nature of any magnetising forces to which the iron has previously been subjected.

Hysteresis.

Thus suppose that, starting with an unmagnetised bar of iron, we gradually increase the magnetising force and determine the corresponding values of the induction, we shall in this way obtain a curve OAC (Fig. 255) similar to the curve given in Fig. 251. If now when the point C is reached the magnetising force is gradually decreased and the value of the induction is again measured as H decreases, it will be found that the curve obtained does not coincide with the curve obtained with increasing values of H, but has the form CD. Thus when the magnetising force is zero, the induction, instead of being zero, has a value OD.

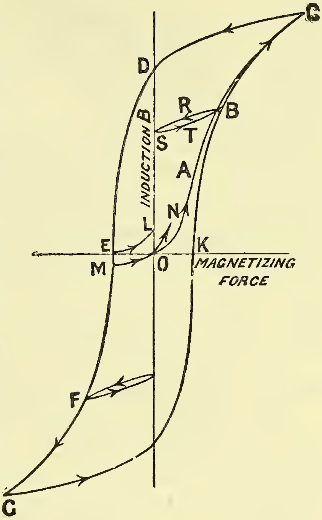


FIG. 255.

If now the direction of the magnetising force is reversed, the curve DEG will be obtained; while on decreasing the magnetising force to zero, and then starting with it in its original direction, the branch GKBC of the curve will be obtained. It will be seen, by a study of this curve, that in all cases the magnetisation appears to lag behind the magnetising force, and to this phenomenon the name *hysteresis* has been applied. If, after the value of H has again reached its maximum positive value, it is again decreased to the same negative value, then back to the extreme positive value, the curve obtained will be very nearly, if not exactly, coincident with the curve CDEGKC. The magnetising force represented by OE or OK represents the force required to deprive the bar of its residual magnetisation. It must, however, be remarked that the condition represented by the points E and K, at which the induction is zero, is quite different from the condition at O, before any magnetising force had been applied. If when the bar is in the condition represented by E the force is reduced to zero, the induction would become positive, and the condition would be represented by the point L. Even if the magnetising field is reversed at M and then decreased to zero, so that the condition of the iron is represented by the point O, where both the force and the induction are zero, the condition of the iron is different from what it was at the start, for if the magnetising force be gradually applied, the B-H curve, as these curves showing the relation between

B and H may be called, is now along ON and not along the original curve OAB.

If, after the iron has reached a condition represented by the point B, the magnetising force is gradually decreased to zero, the curve BRS is obtained, OS as before representing the residual magnetism. If now the magnetising force is gradually increased in the same direction as before, the curve STB will be obtained. Thus in this case also the B-H curve, when the value of H is taken through a cycle of values, encloses a loop. Now work has to be done to increase the induction in a piece of iron, and the greater the existing induction the greater the work that has to be done to increase the induction by a given amount, so that during a cycle a greater magnetising force has to be used to obtain a given induction while we are magnetising the iron than that which corresponds to the same induction when the magnetising field is decreasing. Hence it follows that more work is done during the time that the rod is being taken from G to C than is done by the magnetism of the rod while it is passing along CEG. Thus a certain amount of work has to be done to carry the rod round the cycle represented by the curve, and it can be shown [1] that this amount of work is represented by the area of the loop included by the curve. The energy expended in doing this work appears as heat, which is developed in the iron as a consequence of the changes in its magnetisation.

This hysteresis phenomenon is of very great practical importance, since in dynamos, electric motors, and transformers the magnetism of part of the iron of the machine is being continually reversed, and thus not only is part of the energy supplied to the machine being continually wasted on account of hysteresis, but also the heat developed has to be got rid of or otherwise the machine would be damaged.

The hypothesis that the molecules of iron are magnets and that magnetisation of the iron consists in turning the axis of these molecular magnets parallel to the direction of the magnetising field has been shown by Ewing to explain the characteristics of the magnetisation of iron.

Ewing's molecular magnet.

Thus consider the curve shown in Fig. 256, which gives the relation between the intensity of magnetisation and the magnetising force for a sample of iron. The curve indicates that the magnetisation proceeds in three stages. In the first stage, A, the susceptibility, I/H, is small, the curve starting off at a small inclination to the axis of H. In the second stage, B, the susceptibility increases very rapidly, that is, a small increase of the magnetising force produces a relatively large increase in the induced magnetisation. In the third stage, C, the increase of the intensity of

[1] See Watson's *Text-Book of Physics*, § 509.

magnetisation with increase of H is slow, and for very large values of H is practically nil. There is also a marked difference as regards hysteresis between the sections of the curve. In the first section, on the removal of the magnetising force, the iron loses nearly all its induced magnetism, there being hardly any hysteresis. In the second portion however, on the removal of the magnetising force the iron is able to retain a considerable proportion of its magnetism, while in the third stage the amount of the residual magnetism is hardly greater than in the second stage.

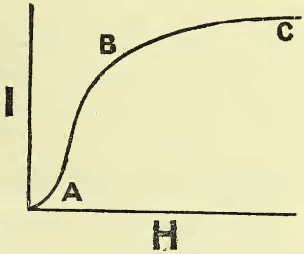


FIG. 256.

As an introduction to Ewing's theory, let us consider the case of two small magnetic needles, which are supported on fixed pivots near each other, but not so near that the poles of the needles can come in contact. If there is no external field these needles will take up a position such as that shown at (a), Fig. 257, in which the axes of the magnets are

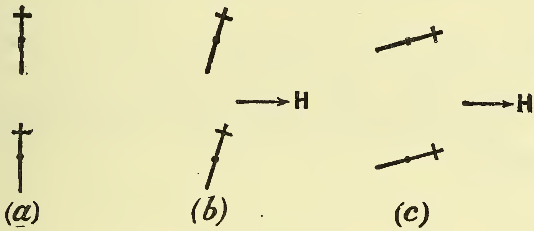


FIG. 257.

parallel to the lines joining the pivots. Suppose now that an external magnetic force, which is at first weak, acts in the direction of the arrow H. As a result of this weak external field the magnets will be slightly deflected, and on the removal of the field they will return to their original positions. This corresponds to the stage A of the I-H curve. If, however, the value of H is increased, a stage will at length be reached when the magnets will suddenly fly round into the positions shown at (c). On further increasing H, the magnets will set themselves more and more nearly with their axes parallel to the direction of the field. If now the value of H is gradually decreased, the inclination of the axes of the magnets to the direction of H will gradually increase, until for some value of H, which will be less than that for which the sudden swing round occurred, the magnets will suddenly return to the position shown at (b). We have, therefore, in this excessively simple arrangement three distinct stages: the first, in which the magnetising

field only produces a small deflection, which is such that on the removal of the magnetising force the deflection becomes zero; secondly, a stage where the magnets reach an unstable position and then suddenly swing round into a new configuration, and where this configuration does not break up until the deflecting force reaches a value smaller than that for which the unstable condition was reached when H was increasing; and, thirdly, a stage when increase of H only produces a small increase in the alignment of the magnets. Thus with only two magnets an indication of the chief peculiarities of the magnetisation curve can be obtained.

By considering much larger numbers of such pivoted magnets a much nearer approach to the phenomena actually found in the case of the magnetisation of a magnetic metal can be obtained. We have, however, said enough to indicate the line of argument by means of which Ewing supports his theory, and for further details we must refer the reader to his original papers on the subject.

In order to account for the heat developed in iron, due to hysteresis, when it is taken through a cycle of magnetisation, Ewing supposes that, on the decrease of the magnetising force, the molecular magnets return towards their undisturbed positions, and in doing so acquire kinetic energy, so that instead of immediately coming to rest they will execute oscillations about their position of rest till the kinetic energy thus acquired is converted into heat due to the currents induced in neighbouring molecules.

CHAPTER IV

ELECTROSTATICS

139. Electrification. Conductors and Non-Conductors.—Thales, who lived about 600 B.C., discovered that amber when rubbed acquires the property of attracting light bodies, such as pieces of pith or cork. Towards the end of the sixteenth century Gilbert showed that this property was also possessed by other bodies, such as wax, sulphur, and glass. All such phenomena are studied in the science of electricity, the name being derived from the Greek name for amber.

A body which has acquired this property of attracting other bodies, the attraction considered being of course different from the gravitational attraction which all bodies exert one on the other, is said to be electrified, or to possess electrification.

The most usual manner of causing the electrification of a body is that referred to above, namely, friction with a suitable rubber. Thus a stick of sealing-wax, when rubbed with a dry piece of flannel, becomes electrified, as also does a rod of glass when rubbed with silk.

All substances may be roughly divided into two classes, called conductors and non-conductors. In a conductor the electrification spreads all over the body, so that if one point of the body is by any means electrified, this electrification immediately spreads all over the body. In the case of a non-conductor, or insulator, as such bodies are also called, the electrification does not spread in this way, but remains in the neighbourhood of the point where the electrification took place.

The best conductors are the metals and solutions of most salts in water, while the best non-conductors are ebonite, glass, shellac, sulphur, paraffin, sealing-wax, and silk. There is, however, no hard and fast line of demarcation between the two classes, for such bodies as dry wood and paper have intermediate properties, and are sometimes called semi-conductors. In the study of electricity it is of much importance to have a good non-conductor, for by this means we are able to support a body in such a way that any electrification communicated to it will not spread to neighbouring bodies through the support. Although no body is known which is a perfect insulator, yet glass, particularly when it has been boiled in water and is then kept in a dry atmosphere, paraffin, and fused quartz are sufficiently good insulators for all practical

purposes. When a body is supported on an insulating stand, we shall speak of it as being insulated.

If a rod of sealing-wax is electrified by rubbing with flannel, and is then suspended by an insulating thread, such as silk, and a second rod of sealing-wax is also electrified in the same way and brought near the first, they will repel each other. In the same way, if two rods of glass are electrified by being rubbed with silk, and one of them is suspended by the silk thread and the other brought near, repulsion will take place. If, however, a rod of sealing-wax, electrified by friction with flannel, is brought near the glass rod, which has been electrified by friction with silk, the two will attract one another. We thus see that we have here to do with two kinds of electrification, in the same way that in the case of magnets we had to do with two kinds of poles. The kind of electrification that is developed in glass when it is rubbed with silk is distinguished by being called positive electrification, while the kind of electrification produced in sealing-wax by friction with flannel is called negative.

We may then state the law of electrical attraction and repulsion as follows: Bodies electrified in the same manner repel one another, while bodies electrified, one positively, and the other negatively, attract one another.

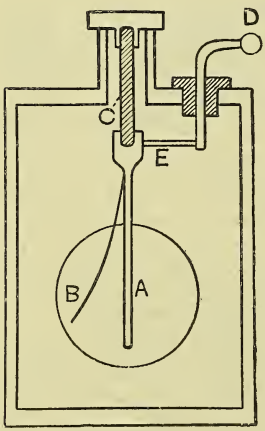


Fig. 258.

Whenever electrification of one kind is produced in any way, electrification of the opposite kind is also produced at the same time. Thus in the case of the glass electrified by friction with silk, while the glass will attract a negatively electrified rod of sealing-wax, the silk used to rub the glass will repel the sealing-wax, thus indicating that the silk has become negatively electrified.

The kind of electrification developed in a body depends on the nature of the body with which it is rubbed; thus while glass becomes positively electrified when it is rubbed with silk, it becomes negatively electrified when it is rubbed with a cat's skin. The kind of electrification produced is also dependent on the state of polish of the surface, on the temperature, &c.

The gold-leaf electroscope.

In order to study the sign, and to a certain extent the magnitude of the electrification produced in a given body, the instrument shown in Fig. 258, and called the gold-leaf electroscope, is often convenient. It consists of a metal box with a window at the back and front. A metal plate A is supported by a rod

of fused quartz or sulphur C, and has attached to its upper end a narrow strip of gold-leaf B. The metal rod D which passes through an insulating plug in the outer case carries a spring E, so that by turning the rod the spring can be brought into contact with the plate A and the electroscope can be charged. When D is put into conducting communication with an electrified body, the spring E being in contact with the plate A, this plate and the gold-leaf B both become charged with the same kind of electrification. Since two bodies charged with the same kind of electrification repel one another, the gold-leaf is deflected, as shown in the figure, the amount of the deflection measuring the amount of the electrification communicated to the electroscope.

If an electrified body is brought near the knob of a gold-leaf electroscope, it will be found that the leaf diverges, showing that the electroscope has become electrified before the electrified body has come into conducting communication with the knob. On the removal of the electrified body the leaf again collapses, showing that it has lost the electrification it possessed when the electrified body was near. This electrification, caused by the *proximity* of a charged body, is said to be produced by induction.

If the inducing body is charged positively, the part of the insulated body nearest to the inducing charge will be negatively electrified, while the part furthest from the inducing charge will be positively electrified. That this is so can easily be shown by means of a small piece of metal attached to an insulating handle, and called a proof-plane, which is brought into contact with different parts of the body on which the induced charges are produced. The sign of the charge carried away by the proof-plane, after contact with any given part of the body, can be found by means of the gold-leaf electroscope. In this way it can be shown that whenever an insulated conductor is placed in the neighbourhood of a charged body, the conductor will become electrified by induction, the electrification at the end nearest the charged body being of the opposite kind to that of the charged body, while the electrification on the end furthest from the charged body is of the same kind as that of the inducing charge. If, while an insulated conductor is in the neighbourhood of a charged body, so that it is charged by induction, it is placed in conducting communication with the earth, the electrification of the same kind as that of the inducing charge will be destroyed. If the connection with earth is now broken, and the inducing charge is then removed, it will be found that the conductor is now electrified with the opposite kind of electrification to that of the inducing body. In the case when the inducing charge is removed before the conductor has been put to earth, the reason why, on the removal of the inducing charge, the conductor was unelectrified was that the two kinds of electrification, produced in equal quantities by the

Electrification by induction.

induction, neutralise each other. The reason the electroscope shown in Fig. 258 is placed in a metal box is to shield the gold-leaf from the direct inductive effect of neighbouring charges, for, as we shall see later, a conducting sheet connected to earth acts as a screen. We shall return to the consideration of the subject of induction after we have dealt with the quantitative measurement of electrification.

140. Coulomb's Law. Unit Charge.—By means of the torsion balance, Coulomb was able to show that the force exerted on one another by two small charged conductors is directly proportional to the product of their charges, and inversely proportional to the square of the distance between the bodies.

Hence, as in the case of the unit magnetic pole, we may define the unit electrification or charge as such that if two small bodies, each charged with a unit, are placed at one centimetre apart in air, the force they will exert on one another will be one dyne. The reason the medium air is specified is that, as we shall see later, the force exerted between two charged bodies depends on the nature of the medium which fills the space between them, while the reason the bodies on which the charges are supposed to exist are taken as small is that if the bodies were of appreciable magnitude the distribution of the electrification would be altered by the action of the one charge on the other.

Suppose then we had two points, charged with e and e' units of electricity respectively, placed at a distance r apart in air, the force, F, which they would exert one on the other, due to their electrification, will be given by the equation

$$F=\frac{ee'}{r^2} \qquad . \quad . \quad . \quad . \quad (156)$$

The force will be an attraction if the charges e and e' are of opposite sign, and a repulsion if they are of the same sign.

The unit charge defined above is called the electrostatic unit of quantity of electricity or charge, to distinguish it from another unit which is often used, and which depends for its value on that of the unit pole. When considering problems involving charges of electricity at rest on conductors the electrostatic unit is generally employed, and in this chapter when a unit of electricity is mentioned we shall throughout mean the electrostatic unit.

141. Electrical Field.—The strength of an electrical field at any point is equal to the force in dynes which a unit positive charge would experience at that point. The strength of an electrical field at any point is often called the *electrical intensity*, or the electrical force at the point.

The space in the neighbourhood of a charged body has, owing to the presence of the charge, properties which it would otherwise not possess.

Thus an uncharged body will in this region become charged by induction, while another charged body will be acted upon by a force. Hence the region near a charged body is said to be an electrical field of force.

If a small body, charged with the unit positive charge, is brought into such an electrical field, this unit charge will be acted upon by an electrical force, which at every point of the field will have a definite magnitude and direction. As in the case of magnetism, a line, such that its direction at every point is the same as the direction of the force acting on the unit charge when placed at the point, is called a line of force. The direction in which a line of force is supposed to run is the direction in which a small positively electrified body would tend to move. Hence a line of force will always start from a body which is positively electrified and end on a body which is negatively electrified.

Electric lines of force.

Just as in the case of a magnetic field of force, we may draw the lines of electrical force in such a way that not only do they indicate the *direction* of the electrical force at the different points of the field, but also that by their closeness they indicate the *magnitude* of the electrical intensity. The convention adopted is, however, different in the case of the electrical field to that adopted for the magnetic field. In the case of the magnetic field the number of lines intersecting unit area perpendicular to the field was taken equal to the strength of the field, and it followed that $4\pi m$ lines originated from a north pole of strength m. In the electrical field the number of lines of force which intersect unit area taken perpendicular to the field is equal to the strength of the field divided by 4π. From this it follows at once, by an argument similar to that used on p. 384, that q lines originate on a body which possesses a charge of q units of positive electricity. This difference between the conventions adopted in the electrical and magnetic fields is liable to cause confusion. Since, however, it is practically universally adopted there seems no likelihood of it being changed in order to secure uniformity.

In the case of magnetism every magnet has both a north and a south pole, and these are of equal strength, so that as many lines of force terminate on a given magnet as leave it. In the case of a charged body we may have the charge all positive or all negative, so that lines of force may only originate on a body or may only terminate on it. A line of force must, however, somewhere originate on a positively charged body and terminate somewhere on a negatively charged body. From each portion of the surface of a charged body on which there exist unit quantity of electricity a single line of force begins or ends. Thus if s lines of force originate from unit surface of a charged body, the quantity of electricity on that unit of surface is s units. The charge on unit surface is called the *surface density* of the electrification.

The lines of force in the case of two small bodies, one of them positively and the other negatively electrified, and placed at a very great distance from all other conductors, so that all the lines of force which leave the positively electrified body terminate on the negatively electrified body, are shown in Fig. 259, while in Fig. 260 the lines of force in the case where the two bodies are electrified with the same kind of electrification are shown.

As in the corresponding case in magnetism, we may account for the attraction which takes place in the one case and the repulsion in the other, if we suppose that there exists a tension along the lines of force, and that something of the nature of an hydrostatic pressure acts at right angles to the direction of the lines, so that they repel one another.

When the electrical charge on any system of conductors alters its

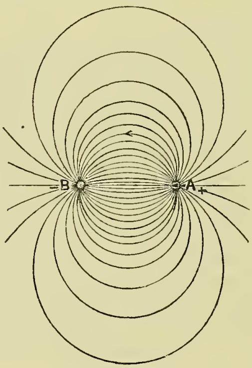


Fig. 259.

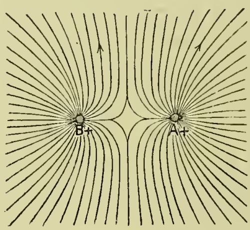


Fig. 260.

distribution, we may consider that each unit of the charge, as it moves over the surface of the conductor, drags the end of its line of force after it, but that, on account of the tension acting along the line, the tendency is for the line to become as short as possible. When the two conductors, on which any given line terminates, are separated by a non-conductor, the line of force cannot shorten indefinitely, for the ends of the line cannot leave the conductors. If, however, the two conductors are placed in conducting communication, say by being joined by a wire, the ends of the line can now move along this wire, so that the line can shorten indefinitely, and ultimately vanish.

Thus by supposing that not only does the tension along the lines of force give rise to a mechanical force acting on the matter on which the electrification exists, but also that this tension causes the electricity of the two opposite kinds which exist at the two ends of the line of force to tend to approach each other, and can only be kept apart by the inter-

position of a non-conductor, we shall be able to explain how it is that one of the kinds of electricity produced by induction remains on the body when the latter is put to earth, while the other kind of electrification escapes to earth.

In Fig. 261 let A represent the inducing body, which we may suppose charged with positive electricity, and B be an insulated conductor which is electrified by induction by A. Then some of the lines of force (shown by the full lines) which leave A will terminate on B, and B will therefore

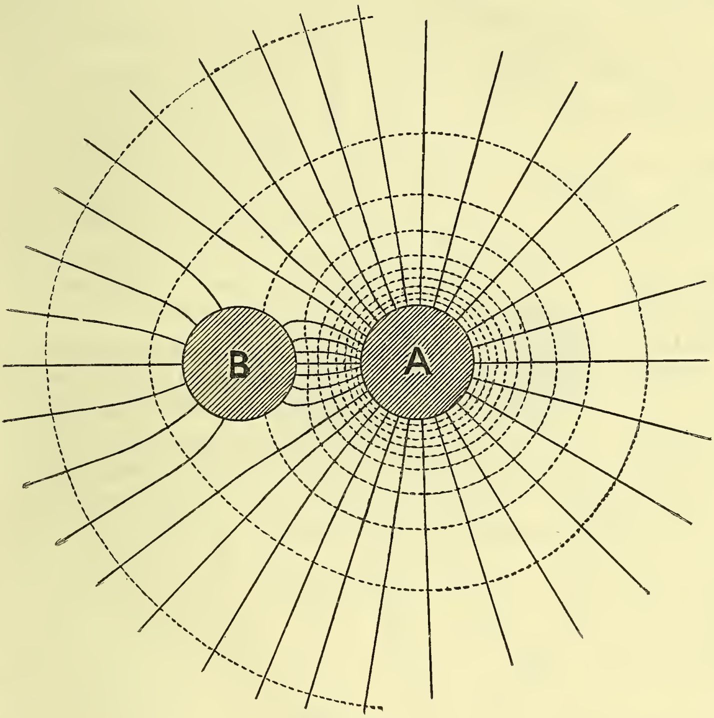


FIG. 261.

be negatively electrified at the part where these lines meet the surface. In addition a number of lines of force will leave B, and terminate on surrounding conductors, such as the walls of the room in which the two bodies are placed. The part of B where these lines leave the surface will be positively electrified, the corresponding negative charge being on the walls. The lines of force which stretch from A to B, by their tension, cause the negative charge on B to accumulate on the side next A. The whole charge does not accumulate at the nearest point however, because of the mutual repulsion which the lines of force exert on one another.

It is owing to this repulsion between the lines that the lines leaving the body B accumulate at the other end.

When the body B is put in conducting communication with the earth, *i.e.* with the bodies on which the lines of force which leave it terminate, owing to the action of the tension on the electrification itself, the latter will escape, but the negative electrification corresponding to the lines of force which leave A and terminate on B will not be able to reach A, since

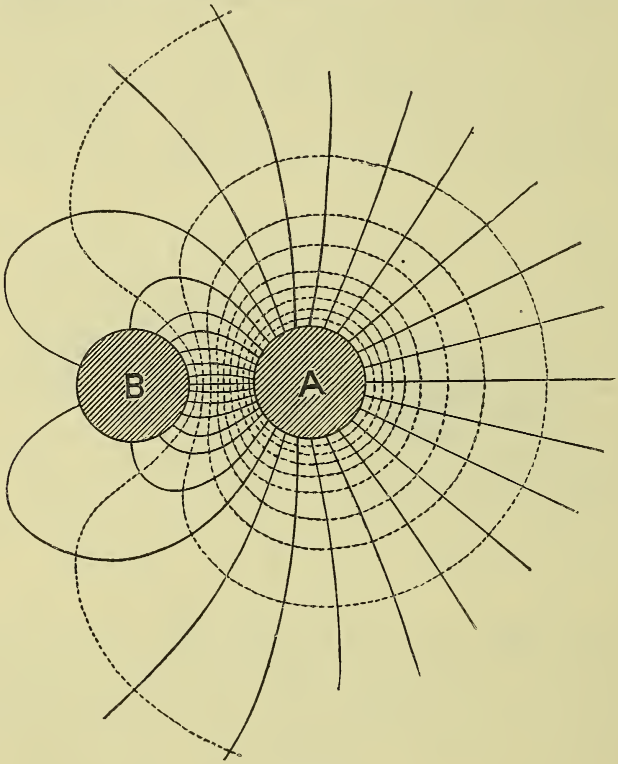


FIG. 262.

these two bodies are not in conducting communication. The distribution of the charges will then be as shown by the lines of force in Fig. 262, where there are no lines of force leaving B, indicating that the charge on B is everywhere negative.

In the case of magnetism the lines of force are continuous and thread through the magnet itself. The electrical lines of force, however, do not extend within the material of a *conductor*. The proof of this is the fact that *within* a hollow conductor there is no electrical field, however strongly

the conductor is electrified. Further, the charge on a conductor is confined entirely to the outside surface.

Electrical charge confined to outside surface of a conductor.

The truth of the above statements can be demonstrated by taking a hollow metal vessel supported on an insulating stand, a hole in the top of the vessel allowing of the introduction of a small charged sphere supported on an insulating handle. If the charged sphere is introduced and allowed to touch the inside of the vessel, and then removed, taking care that it does not touch the edge of the hole in the vessel, the sphere will be found to have *entirely* lost its charge whether the vessel is originally uncharged or has a large charge. When touching the inside of the vessel the sphere formed part of that conductor, and the whole charge immediately travelled to the outside.

Equal quantities of + and − electricity produced by induction.

The above arrangement will also enable us to study the induced charge produced by the small charged sphere. If the vessel is uncharged and a charge $+q$ is introduced on the insulated sphere, the sphere *not* being allowed to touch the vessel, an induced charge will be produced on the vessel which is positive on the outside and negative on the inside. The production of these induced charges is conveniently studied by considering the lines of force. Before its introduction into the vessel the q lines of force which leave the sphere terminated on surrounding conductors and the walls of the room. When the sphere is introduced within the vessel each line of force is divided by the walls of the vessel into two parts—first, a part stretching from the sphere to the inner wall of the vessel, and second, a part stretching from the outer wall of the vessel to the surrounding objects and the walls. Since q lines leave the sphere, q lines must terminate on the inner wall of the vessel, and hence a charge of $-q$ units must exist on this inner surface. Similarly q lines leave the outer surface of the vessel, and hence there must be a charge of $+q$ units on this outside surface.

That the induced charges are $-q$ and $+q$ respectively on the inner and outer surfaces can be shown by having the vessel connected to a gold-leaf electroscope, and noting the deflection on the introduction of the charge. Next let the vessel be connected to earth so that the charge on the *outside* escapes and the electroscope leaf collapses. On carefully removing the charged sphere, without allowing it to touch the vessel, the electroscope leaf will be deflected to the same amount as before, showing that the charge on the inside is equal in magnitude to the charge previously induced on the outside. If the vessel is now discharged, and the charged sphere again introduced and allowed to touch the inside of the vessel, the deflection of the electroscope will be the same as that previously obtained, showing that the positive and negative charges

which were induced were each numerically equal to the inducing charge. The above experiment was first performed by Faraday, using an ice-pail for the hollow vessel, and hence it is generally called Faraday's ice-pail experiment.

142. Difference of Potential.—If two conductors, one of which is charged positively and the other negatively, are put in conducting communication, their state of electrification will become changed, so that if they originally possessed equal charges they will both, after being connected, exhibit no signs of electrification. If the charge on one was greater than that on the other, then, after being connected, the sign of the charge on the two will be the same as the sign of the charge which was originally the greater, while the sum of the charges now possessed will be equal to the difference of the two original charges. If, however, two bodies, each of which is charged with electricity of the same kind, are put in conducting communication, it does not follow that the charge on the body which was originally electrified with the larger charge will be decreased and that of the other increased, for a small sphere charged with one unit of positive electricity, when put in conducting communication with a sphere, of which the radius is three times that of the other and which is charged with two units of positive electricity, will lose electrification. There must evidently, therefore, be some other condition besides the magnitude of the charge which decides whether, when two charged bodies are put in communication, the charge of one or other of them becomes increased.

Two conductors are said to be at different potentials if, when they are put in conducting communication, the distribution of electrification on the conductors changes. The body on which the positive electricity *decreases* is said to be at the higher potential.

This idea of electrical potential is of the same nature as the idea of temperature in the case of heat, or of level in the case of the flow of water in a pipe, for, as we have seen, heat always flows from a body at a higher temperature to a body at a lower temperature, and water only flows from places at a higher level to places at a lower level.

The difference in potential between two charged conductors is measured by the work that would have to be done on a small body charged with a unit of positive electricity when the body is moved from the immediate neighbourhood of the conductor at the lower potential to the immediate neighbourhood of the conductor at the higher potential.

For all practical purposes the measure of the available energy of a waterfall is known if the available head and the quantity of water which passes in a second are known, for the variation in the value of the acceleration due to gravity (g) is comparatively small. It would be quite otherwise, however, if the value of g varied to any great extent from one

place to another on the surface of the earth. Thus suppose that we had to do with two waterfalls in which the quantity of water which passed per second was the same, but the fall was different and the value of g was twice as great at one place as at the other. Then the work which could be obtained from the unit mass of water as it passed from the top to the bottom of the fall would be gh_1 in the one case and $2gh_2$ in the other. Hence, as far as the quantity of energy available is concerned, the height through which the water falls, that is, the difference in level between the water above and below the fall, is not a measure of the value of the fall. If, however, we measured this difference of "level" by the quantity of work which must be done to raise unit mass of the water from the bottom of the fall to the top, then the available energy of any fall would be obtained by simply multiplying this quantity by the quantity of water which passes over the fall in a unit of time. Now although, as has been mentioned above, the changes in g are so small as to make it quite unnecessary to adopt, in the case of waterfalls, any such device, yet it will be seen why the method adopted for measuring the difference of potential between two charged bodies is quite a reasonable one.

When considering the absolute scale of temperature in § 81 we used a very similar method, for the difference in temperature between two bodies was measured by the *work* which could be done by a reversible engine when working between these two temperatures, and taking a given quantity of heat from the hotter body. Thus in this case also a quantity of work is used as a measure of the difference of the quantity (temperature) which decides in which direction heat will flow when two bodies are placed in thermal communication, and is, therefore, analogous to potential in the electrical problem.

The amount of work done on the unit of positive electricity as it is carried from the neighbourhood of one charged body to that of the other is the same, whatever the path by which it is moved. If it were not, so that it were possible to pass from a point A (Fig. 263) to another point B at a lower potential, in such a way that the work w_1 done on the unit charge when taken along the path ACB was greater than the work w_2 done when the unit is moved along the path ADB, then by taking the body with the unit charge from A to B by the path ACB and bringing it back by the path BDA, the whole system would have performed a cycle, for the initial and final states are the same, while an amount of work equal to $w_1 - w_2$ would have been done without the supply of any external energy. This being contrary to the doctrine of the conservation of energy, it follows that w_1 must be equal to w_2, that is, the work done when the

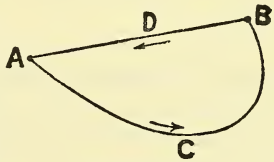


Fig. 263.

unit charge is carried from A to B must be independent of the path by which it is carried.

We have defined the difference between the potential of two points and shown how it is measured, and we have now to choose some fixed potential as the zero of potential. The potential of the earth is usually taken as the zero of potential, so that the potential of a positively electrified body is positive, and that of a negatively electrified body is negative, for a positively electrified body will repel a body charged with a unit of positive electricity, and so work will be done on the unit charge as it is moved from the electrified body to the earth, while if the body is negatively electrified, work must be supplied to move the unit charge from the electrified body to the earth.

Equipotential surfaces.

An *equipotential surface* is a surface such that the potential of all points upon it is the same. No work is therefore done when a charged body is moved along a path which lies on an equipotential surface. It follows at once that the lines of force must always cut an equipotential surface at right angles. If a line of force did not cut an equipotential surface at right angles, then the force which acts in the direction of the line of force can be resolved into two components, one along the surface and the other normal to the surface. If an electrified body were placed at the point where the line of force cuts the equipotential surface, it would be acted upon by the component parallel to the surface, and if it were moved in the direction in which this component acts, work would either be done on or by the electrified particle. But by the definition of an equipotential surface no work is done when a charged body is moved from one point of such a surface to any other point on the surface. Hence it follows that the component of the force parallel to the surface of the equipotential surface must be zero, or, in other words, that the direction of the line of force must be perpendicular to the surface at the point where it cuts the surface.

Since in the case of a conductor the electrification is not prevented from spreading itself over the surface of the body, no change in the distribution of the electrification would take place by connecting any two points of the surface by a conducting wire, and so all parts of the surface must be at the same potential. The surface of a conductor must therefore be an equipotential surface, and hence the lines of force must always cut the surface of a conductor at right angles to the surface.

In Fig. 264 the lines of force and the equipotential surfaces for a positively charged body A are shown, the traces of the equipotential surfaces being shown by the dotted lines. If an insulated uncharged conductor B is placed in the neighbourhood of the charged conductor,

this conductor will become electrified by induction. Now if the conductor B could be brought near the charged body A without producing any change in the distribution of the charge on the conductor or changing the state of the electrical field in the space now occupied by the conductor, that is, if the lines of force and the equipotential surfaces were to remain as in Fig. 264 *after* the introduction of the conductor, then those parts of the conductor B furthest from A would be at a lower potential than the parts nearer A. Hence, since it is impossible for different parts of a conductor to be at different potentials so long as the electrification is not changing, some change

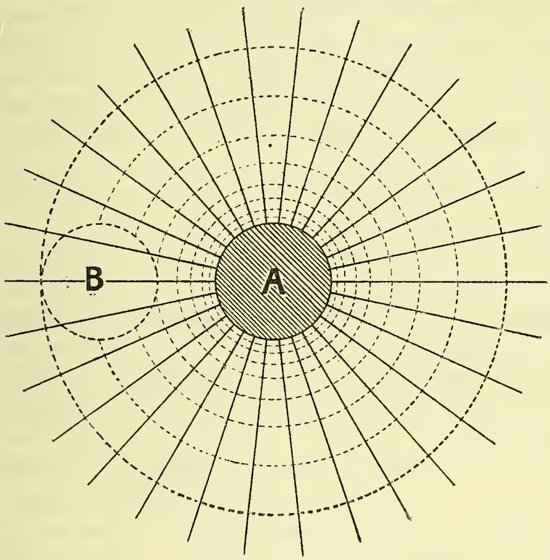


FIG. 264.

in the electrical conditions must take place so as to raise the potential of the more distant parts of the conductor B, or lower the potential of the nearer parts. This will occur if the more distant parts become positively electrified and the nearer parts negatively electrified, for under these conditions a greater repulsive action will be exerted on a unit of positive electricity when placed near to the further surface of B, and hence a greater amount of work will be done on this unit while it is being moved from this position to the neighbourhood of the earth. In the same way, less work will be done when carrying a unit charge from the near side to the neighbourhood of the earth, so that the potential of the near side will be reduced

by the presence of the induced negative electrification. This lowering of the potential on the near side of B itself involves a lowering of the potential of the near side of the conductor A, and hence also of the far side. This lowering of the potential of the far side is produced by the accumulation of the positive electrification of A on the side near B.

The form of the lines of force and of the equipotential surfaces under the new conditions is shown in Fig. 261. It will be seen that the change in the distribution of the charge on A, as well as the distribution of the induced charge on B, is such that the surfaces of the two conductors are equipotential surfaces. If the insulated conductor is earthed, then the electrification on both conductors is altered, but in such a way that, as shown in Fig. 262, the surfaces of the conductors remain equipotential surfaces. We thus see that the fact that the distribution of the electrification on the body, when placed in the neighbourhood of a charged body, is not uniform, is not inconsistent with the surface of the conductor being an equipotential surface, but is in fact the distribution which, in conjunction with the inducing charge, ensures the fulfilment of this condition.

If A and B are two points on a line of force, they must necessarily be at different potentials. Let the potential of A be V_1, and that of B be V_2 (V_1 being greater than V_2); then if a small body carrying the unit charge of positive electricity is moved along the line of force from A to B, the work done will be equal to $V_1 - V_2$, for the difference in potential between two points is measured by the work done on the unit charge when it is moved from one point to the other. If the points A and B are very close together, the force F, which acts on the unit charge in the direction from A to B, as it is moved from A to B, may be supposed to remain constant, and to be equal to the average force that acts. The work done by the unit charge as it is moved from A to B will therefore be equal to Fs, where s is the distance from A to B, measured along the line of force, that is, the distance through which the charge is moved along the line of action of the force F. Hence the work done *on* the unit charge is $-Fs$. Equating the two expressions we have now obtained for the work done on the unit charge when it is moved from A to B, we get

$$V_1 - V_2 = -Fs,$$

or

$$F = -\frac{V_1 - V_2}{s} \quad . \quad . \quad . \quad . \quad (157)$$

Now the expression on the right-hand side of this equation is the difference of potential between the two points divided by the distance between the points measured along a line of force, or, in other words, is

the rate of increase in the potential along the line of force at the points A and B, which are by supposition very close together. Hence the force which acts on a unit charge of positive electricity, when placed in an electrical field, is equal to minus the rate of increase of the potential along the line of force at the given point. If the force acting on the unit charge is constant, it follows that the rate of change of the potential must also be constant. Hence in a uniform electrical field the rate of change of the potential in the direction of the lines of force must be constant, and thus the length of a line of force intercepted between two consecutive equipotential planes will be the same, if the difference of potential between consecutive equipotential planes is itself constant.

Force exerted on a charged body within a hollow charged conductor.

Since the surface of a charged conductor is an equipotential surface, the whole of the space within must be at the same potential, so long as there are no charged bodies within the conductor. For, suppose that within the surface of the conductor there were an equipotential surface corresponding to a higher potential than the potential of the surface of the conductor, then there would be lines of force running everywhere to the outer equipotential surface from this inner one; and since these lines of force must of necessity start from a positively electrified body, it would follow that there must be a positively electrified body within the conductor, which is contrary to our original supposition. In the same way it would follow that, if there existed an equipotential surface of lower potential than that of the surface of the conductor, there must be a negatively electrified body within the conductor. We are therefore led to the conclusion that there can be no point within a closed conductor at a different potential from that of the surface, unless there are charged bodies within the conductor.

Since the strength of an electrical field is equal to minus the rate of change of the potential, it follows that if the potential is constant, that is, if its rate of change is zero, there will be no electrical force exerted within the conductor. We thus see that it follows, from the fact that the charge of a conductor is confined to the outside surface, that there is no force exerted within a charged conductor; and it can be shown that this condition can only be fulfilled if Coulomb's law (§ 140), that the force exerted between two charged bodies varies inversely as the square of the distance, is true.

This method of proof of the accuracy of Coulomb's law admits of a much more vigorous experimental test than the original torsion-balance experiments by means of which Coulomb first proved his law.

Suppose a conducting sphere of radius r, which is at a great distance from all other conductors, has a positive charge of Q units, then the lines of force will be equally spaced radial lines, and the number of

lines which leave unit area of the surface of the sphere will be $Q/4\pi r^2$.

Intensity due to a uniformly charged sphere.

Hence since the electrical intensity is equal to 4π times the number of lines per unit area we have that the electrical intensity at a point just outside the sphere is equal to Q/r^2. The surface density, s, of the charge is $Q/4\pi r^2$. Hence the intensity just outside the surface is equal to $4\pi s$. If we draw a concentric sphere of radius R, the number of lines of force which intersect unit area of this sphere is $Q/4\pi R^2$ and the electrical intensity is equal to Q/R^2. It will thus be seen that the electrical intensity at any point outside the charged sphere is the same as it would be suppose the whole charge were concentrated at the centre of the sphere. Hence it also follows that the potential at any point outside the sphere is the same as that which would be produced if the whole charge were concentrated at the centre of the sphere. The potential at all points *within* the sphere will be equal to the potential at the *surface* of the sphere, that is equal to the potential at a distance r from a point charge of Q units.

We have now to calculate what will be the potential at a distance d from a point charge of Q units. Let us draw a line of force through

Potential at a given point due to a small charged body.

the given point P. This line will be a straight line passing through the charge, and the force acting on a unit positive charge placed at P will be Q/d^2. If now starting from P we carry the unit charge along the line of force to infinity the work done will be equal to the potential at P. Suppose the path divided into a number of *small* elements PP_1, P_1P_2, P_2P_3, &c., the distances of P_1, P_2, P_3, &c., from the charge being d_1, d_2, d_3, &c. The force acting on the unit charge when at P_1 will be $Q/d_1{}^2$, and the *mean* force acting as the unit charge is carried from P to P_1 will be somewhere intermediate between Q/d^2 and $Q/d_1{}^2$. If P_1 and P are very near together the mean value of the force will be very nearly equal to the geometrical mean of these two quantities, namely Q/dd_1. The distance traversed between P and P_1 is d_1-d, and hence the work done is

$$\frac{Q(d_1-d)}{dd_1}, \text{ or } Q\left\{\frac{1}{d}-\frac{1}{d_1}\right\}$$

In the same way the work done when the charge is carried from P_1 to P_2 is

$$Q\left\{\frac{1}{d_1}-\frac{1}{d_2}\right\}$$

and so on. Hence the total work done when the unit charge is carried to infinity, that is the potential V at P, is given by

$$V=Q\left\{\frac{1}{d}-\frac{1}{d_1}+\frac{1}{d_1}-\frac{1}{d_2}+\&\text{c.}-\frac{1}{\infty}\right\}$$

or

$$V=\frac{Q}{d} \quad . \quad . \quad . \quad . \quad . \quad (158)$$

Hence the potential at a distance d from the centre of a sphere charged with Q units is equal to the charge divided by d. If the sphere has a radius r, the potential at a point on the surface of the sphere, *i.e.* at a distance of r from the centre, is Q/r.

143. Capacity.—We have seen in the last section that the potential V of a sphere of radius r when charged with Q units is Q/r. Hence, since Q/V is equal to r, which is a constant for the given sphere, we see that the ratio of the charge to the potential to which it raises the sphere is a constant. This result is quite general in that it is found that whatever the shape of the charged conductor the ratio of the charge to the potential to which this charge raises the conductor is a constant for the particular body. This constant is called the *capacity* of the conductor. Hence if a charge Q raises the potential of a conductor to V, the capacity, C, is given by

$$C = Q/V \quad . \quad . \quad . \quad . \quad (159)$$

The capacity may therefore be defined as the charge required to raise its potential by unity.

From what has been said above it will be seen that the capacity of a sphere is numerically equal to its radius measured in centimetres.

Condensers.

We have seen on p. 422 that if an uninsulated conductor is brought near a charged body, the potential of this latter is diminished on account of the induced charge on the uninsulated conductor. Hence the potential of the insulated conductor produced by a given charge is less when the uninsulated conductor is near than it is when this conductor is absent; in other words, the effect of bringing the uninsulated conductor near the charged one is to increase the capacity of this latter.

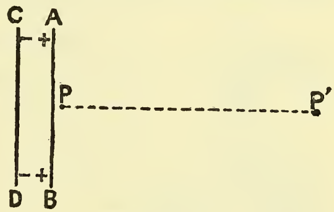


Fig. 265.

We may consider the same problem in a somewhat more direct way, if we suppose that a given conductor, say a plane AB (Fig. 265), is insulated and then charged to a potential V when at a distance from all other conductors.

Let a second plane, which is connected with earth, be placed at such a distance from AB that its presence does not appreciably affect the electrical condition of AB. Then the work that is done in carrying a unit of positive electricity from a point P near AB to a point P′, which is at zero potential, is equal to V. Next suppose that the uninsulated plane is moved near to AB, into the position CD, so that an appreciable charge is induced on it. The work that will now be done while carrying the unit charge

from P to P′ by the same path as before will be less than before, for, on account of the attraction exerted on the unit charge by the negative charge induced on CD, the force exerted on the unit is everywhere less than it was before. Hence the potential of AB is less than it was before. As the plane CD is moved nearer to AB the amount of the induced negative charge increases, and the influence of this negative induced charge in diminishing the repulsive force exerted on the unit charge becomes greater and greater, and hence the potential of AB becomes less and less. The charge on AB remains however the same, and therefore, since the potential to which this charge is capable of raising AB diminishes as the uninsulated conductor CD is brought near, it follows that the capacity of AB must increase as the conductor CD is brought near. If, instead of keeping the charge on AB constant, we had kept the potential constant, then we should have had to increase the charge on AB as the conductor CD was brought up.

An arrangement of two conductors, one of which is insulated and the other uninsulated, placed near one another with an insulator between, is called a condenser. The name condenser was given to such an arrangement on account of the fact that the presence of the second uninsulated conductor appears to exert a condensing action on the electrical charge on the insulated conductor, so that for a given potential it can receive a much greater charge than it could without the presence of the uninsulated conductor.

The capacity of a condenser is the charge which must be communicated to the insulated conductor to raise its potential through one unit of potential. The two conductors of a condenser are sometimes called the armatures of the condenser.

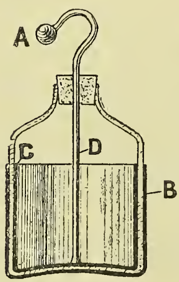


Fig. 266.

The commonest form of condenser is that shown in Fig. 266, and is called a Leyden jar. It consists of a glass jar, the interior of which is coated with tinfoil up to within an inch or so of the top, and a metal knob which is in conducting communication with this inside coating. This tinfoil forms the insulated armature of the condenser, the uninsulated armature being formed by a coating of tinfoil on the outside of the jar.

Another form of condenser which is commonly used consists of a plate of glass, mica, or some other insulating material, which is coated on each side with a sheet of tinfoil or some other conductor, a margin of an inch or so being allowed all round the edge of the glass. One coating is connected with earth, and the other forms the insulated armature of the condenser.

If the insulated armature of a condenser is charged to a potential of

V, the other armature being at a potential zero, and this armature is then insulated, while the armature which was at first insulated is put to earth, this armature will not lose much of its charge, as now the rôles of the two armatures are reversed, for what was originally the induced charge is now the inducing charge, while the former inducing charge is now the induced charge. When a condenser is charged, most of the lines of force stretch across from one armature to the other, few stretching from the insulated armature to surrounding objects. Now, in order to discharge a charged conductor, the bodies on which the other ends of the lines of force which leave the conductor terminate must be put in conducting communication with the conductor. For we may imagine that when a charged body is put to earth by means of a conducting communication, such as a wire, that the two ends of each line of force travel along the conducting wire towards one another, the line of force shortening up in virtue of the tension which exists along every such line, until the two ends come together and the line of force shrinks to nothing. In the case of the condenser, if the two armatures are put in conducting communication all the lines of force are able to shrink to nothing, that is, the condenser becomes completely discharged. If, however, after charging the uninsulated armature is insulated, and the other armature is put in conducting communication with earth, only those lines of force which stretch from this armature to the surrounding uninsulated conductors, such as the walls of the room, will be able to shrink and vanish. The great majority of the lines which stretch from one armature to the other will not be able to shrink, for the armatures are not in conducting communication.

Capacity of a spherical condenser.

The calculation of the capacity of a condenser from its dimensions is in general difficult; the case, however, of a condenser consisting of two concentric spheres with air between can easily be solved. Let the inner sphere of radius r be insulated, and the outer spherical shell, of which the *internal* radius is R, be connected to earth. If the charge on the inside sphere is $+Q$ units, as we have seen on p. 417, the induced charge on the inner surface of the outer sphere is $-Q$ units. If the charge on the outer sphere were alone present the potential throughout the space inside would be the same and equal to $-Q/R$. Hence the potential at the surface of the *inner* sphere due to the induced charge on the *outer* sphere is $-Q/R$. The potential at the surface of the inner sphere due to its *own* charge is Q/r. Hence, as the potential at a given point due to the simultaneous action of two charges is equal to the sum of the potentials due to each charge separately, the potential, V, of the inside sphere is given by

$$V=\frac{Q}{r}-\frac{Q}{R}=\frac{QrR}{R-r}$$

Hence if C is the capacity of the condenser,

$$C=\frac{rR}{R-r} \quad . \quad . \quad . \quad . \quad (160)$$

If the radii r and R are very nearly equal, and the thickness of the layer of air between the plates of the condenser is called d, $C=R^2/d$. If S is the area of a sphere of radius R, the capacity can be expressed in the form

$$C=\frac{S}{4\pi d} \quad . \quad . \quad . \quad . \quad (161)$$

Although this expression is strictly only applicable to spherical condensers, yet it holds approximately whenever we are dealing with a condenser in which the distance between the plates is small compared to the area of the plates, whatever be the shape of the plates.

144. Specific Inductive Capacity.—We have hitherto supposed that the medium surrounding the charged bodies has been air, and the expressions we have obtained for the intensity, potential, capacity, &c., refer to air as the dielectric. Now if we have a condenser, say a spherical one such as that described on p. 427, and replace the air between the armatures by another dielectric, such as paraffin, it is found that the capacity of the condenser is altered. The ratio of the capacity of the condenser with a given material as dielectric to the capacity of the same condenser with air as a dielectric is, however, the same whatever the shape or size of the condenser. Hence this ratio measures a distinctive property of the dielectric, and is known as its *specific inductive capacity or dielectric constant.* Hence if the capacity of a condenser is C when it has air as the dielectric, on replacing the air by a medium of specific inductive capacity K, the capacity will become KC. It will be noticed that the definition of specific inductive capacity given above assumes that the specific inductive capacity of air is unity.

If we have a spherical condenser of capacity C and charge it with Q units when the dielectric is air, the potential of the insulated armature will be Q/C, and hence as the potential of the earthed armature is zero the work which would be performed in taking a unit charge from the neighbourhood of the surface of the inner (charged) sphere to the neighbourhood of the inner surface of the outer (earthed) sphere is Q/C. If now we replace the air by a dielectric of specific inductive capacity K, and again charge the condenser with Q units, the difference of potential between the two spheres is Q/KC, and the work which is done when unit charge is taken from one armature to the other is $1/K$ of the former value. As the path over which the unit is moved is the same in the two cases, we infer that the force acting on the unit charge with the dielectric of specific inductive capacity K is $1/K.th$ of the force

which acts when the dielectric is air. In other words, the electric intensity is inversely proportional to the specific inductive capacity. Thus the electric intensity at a distance d from a sphere charged with Q units when the sphere is surrounded by a medium of specific inductive capacity K is Q/Kd^2. Similarly the potential is Q/Kd. It also follows that the force exerted between two charges q and q' when at a distance d apart in the dielectric is qq'/Kd^2.

The dielectric evidently plays a most important part when we are considering electrical phenomena, and an experiment due to Kerr shows that in an electrical field the dielectric is in a state of strain. Kerr found that the dielectric between the plates of a charged condenser behaves as if it were double refracting (see p. 376), although when the condenser is discharged no such effect is observable. Now it is also found that a piece of well annealed glass ordinarily shows no signs of double refraction, but that if it is strained (see p. 377), say by being slightly bent, it immediately becomes doubly refracting. Hence we infer that the dielectric when in an electrical field must be strained in some way. This view is also supported by the fact that in the case of a condenser with a solid dielectric, such as glass, if we give it a charge, then discharge it by momentarily connecting the armatures, after a little time a further charge, called the residual charge, will be found on the condenser. Now, when a solid body is strained it is often found that it only recovers from the strain very slowly, so that in the residual charge of a condenser we have a similar phenomenon in the case of the electrical strain set up when the condenser is charged.

Dielectric in an electric field is in a state of strain.

It is probable that the energy possessed by a charged body (see next section) is stored up as strain of the dielectric, just as the energy of a coiled spring is stored up due to the strain of the steel of the spring.

145. Energy of a Charged Condenser.—Suppose that a condenser of capacity C is charged to a potential V, the uninsulated armature being at the potential zero. Since the potential of the one armature is V, and that of the other is O, the work done in moving a unit charge from one armature to the other will be V.

Let us suppose the condenser to be discharged by carrying the charge from one plate to the other, the quantity e taken at each journey being small. After the first journey the charge on the condenser is $VC-e$, and hence the difference of potential between the armatures is $V-e/C$. After the second journey the difference of potential is $V-2e/C$, and so on. While after the nth journey, where $ne=VC$ or $n=VC/e$, the condenser will be completely discharged.

Now, if the difference of potential between the armatures of the condenser remained exactly V during the whole time the first charge e was

being transported, the work done during the journey would be Ve, while if the difference of potential during the second journey remained at the value it had after the loss of the first charge e, *i.e.* at $V-e/C$, the work done would be $(V-e/C)e$, and so on. Hence the total amount of work done in discharging the condenser would be $eV+eV-\frac{e^2}{C}+eV-\frac{2e^2}{C}+\ldots+eV-\frac{n-1}{C}e^2$. This is an arithmetical progression of which the last term $eV-(n-1)e^2/C$ is equal to e/C, for $n=VC/e$. The sum of the terms is equal to half the sum of the first and last terms multiplied by the number of terms; that is, the total work done is $\frac{1}{2}VC(V+e/C)$.

Now, this result is obviously too large, since when we commence to move the charge e from one armature to the other, the potential between the armature begins to fall, and hence the work done is less than it would be if the potential remained constant and equal to its value at the commencement of the journey. On the other hand, if we suppose that throughout each journey the potential remains constant and equal to the value it has at the *end* of the journey, the work calculated on this assumption will be too small. In this case the work done is

$$eV-\frac{e^2}{C}+eV-\frac{2e^2}{C}+eV-\frac{3e^2}{C}+\ldots+eV-\frac{ne^2}{C};\text{ or }\frac{1}{2}VC(V-e/C).$$

The true value of the work must be intermediate between the values we have obtained on the two suppositions, and we see that when e is very small the two expressions become the same, so that the work done in the discharge of the condenser must be $\frac{1}{2}V^2C$. This expression may also be written in the forms $\frac{1}{2}QV$ and $\frac{1}{2}\frac{Q^2}{C}$, where Q is the original charge of the condenser. Since the work done in the discharge must be equal to the work done during the charge, the above expressions also express the work done in charging a condenser; in fact, these expressions give the energy of a charged condenser due to the charge. We may look upon a charged condenser as possessing stored-up energy due to the strain which is set up in the dielectric, just as the coiled-up spring of a watch possesses energy due to the state of strain it is in owing to its deformation. The spark and the accompanying noise on the discharge are both evidence of the energy which is set free when a condenser is discharged.

Suppose we have a condenser of capacity C_1 and give it a charge Q, it will then possess an amount $Q^2/2C_1$ of electrical energy. If now it is connected to a second uncharged condenser of capacity C_2, so that the two share the charge, the combined capacities are C_1+C_2 while the charge is Q, hence the energy is $Q/2(C_1+C_2)$. Since C_1+C_2 is greater than C_1, the energy of the two condensers is less than that of the first

before it shared its charge. The electrical energy which has thus disappeared is represented by the energy of the spark which always passes when the uncharged condenser is connected to the charged one.

If we have a condenser consisting of two parallel plates with air as a dielectric, and give it a charge, then on separating the plates the capacity of the condenser will decrease, and hence, since the charge is constant, the electrical energy will increase. This increase of energy is supplied by the work which has to be performed in separating the plates against the electrical attraction.

Since the dielectric in an electric field is in a state of strain, we may look upon the energy of a charged body as being stored up in the dielectric, and it can be shown that if we suppose that at a point of an electrical field where the electric intensity is F there is stored $F^2/8\pi$ ergs of energy per cubic centimetre, then the electrical energy of the charged body or bodies to which the field is due will be equal to the sum of this stored energy taken throughout the electrical field.

146. Electrostatic Measuring Instruments.—Suppose that two conducting planes, AB and CD (Fig. 267), are placed parallel to one another at a distance d apart, and the upper plate is charged to a potential V, the lower plate being earthed and hence at zero potential. The lines of force starting from the upper plate will mostly stretch across to the lower, and although near the edges the lines will curve outwards as shown in the figure, near the centre of the plates the lines will stretch straight across from one plate to the other, that is, we shall have a uniform field between the central parts. Let us now consider a portion E of the upper plate of area S, and we will suppose that the surface density of the charge on E is s, so that s lines of force leave each unit of area of E, s lines terminate on each unit of area of the portion of AB which is opposite to E, and s lines will cross unit area at right angles to the lines taken anywhere between. Hence (p. 424) the electrical intensity, F, between the central parts of the two plates is equal to $4\pi s$, and the work which would be performed when moving a unit charge from one plate to the other is Fd, or $4\pi sd$. Hence the difference of potential between the plates is given by

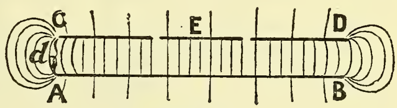


Fig. 267.

$$V = 4\pi sd \quad . \quad . \quad . \quad . \quad . \quad (a)$$

The charge on the part E is sS, and hence the energy of this charge is $\frac{1}{2}sSV$, or $2\pi s^2Sd$.

Next imagine that we allow CD to move in towards AB through a small distance x, the potential will now be given by $V_1 = 4\pi s(d-x)$, and

since the charge on E remains the same, the electrical energy of E is $2\pi s^2 S(d-x)$.

Now, if f is the force acting on E due to the electrical attraction of AB, the work performed when E moves through a distance x is fx, and this must be equal to the decrease in the electrical energy. Hence

$$fx = 2\pi s^2 Sd - 2\pi s^2 S(d-x)$$

or

$$f = 2\pi s^2 S.$$

From (a) above we have $s = V/4\pi d$. Hence

$$f = \frac{V^2 S}{8\pi d^2}$$

or

$$V = \sqrt{\frac{8\pi d^2 f}{S}} \quad . \quad . \quad . \quad (162)$$

Thus if we measure the force exerted on the portion E of the plate CD, when the distance between the plates is d, we can obtain from the above expression the value of the difference of potential V. Thus V is determined in terms of the units of force and length independent of the value of any electrical units.

The portion of the plate CD, which surrounds the part E on which the attractive force is measured, is called by Lord Kelvin, to whom the arrangement is due, the guard ring. The functions of the guard ring are simply to ensure that the electrical field at the part of the plates where the attracted part E is placed shall be uniform.

In the instrument depending on this principle invented by Lord Kelvin, and called the attracted disc electrometer, or the absolute electrometer, the part E on which the force is measured consists of a metal disc supported by three springs, so that it lies concentrically within a circular hole in the guard ring, to which it is electrically connected. The springs are so arranged that when the attracted disc is attracted with a certain force by the opposite plate, AB, it lies exactly in the plane of the guard ring, as indicated by means of two sights which are attached. The plate AB can be moved in a direction parallel to its normal by means of a micrometer screw. When using the instrument the guard ring and attracted disc are connected with the earth, so that their potential is zero, and the other plate is connected with the body of which the potential is to be measured. The distance between the two plates is then altered till the disc comes into its sighted position. The force necessary to bring the disc into its sighted position is determined once for all by placing weights on it, and hence, knowing this quantity (f in the formula), and also knowing the distance between the plates from the reading of the micrometer screw, the potential can be obtained.

The attracted disc electrometer is not suitable for rapidly making measurements of difference of potential, and further it is only suitable for measuring fairly large differences of potential. A more sensitive and convenient instrument, also invented by Kelvin, is the quadrant electrometer. One form of this instrument is shown in Fig. 268. A very light metal needle N is suspended by a fine wire so that it hangs inside a conducting-box. This box consists of four quadrants, C, C′, D, D′, which are

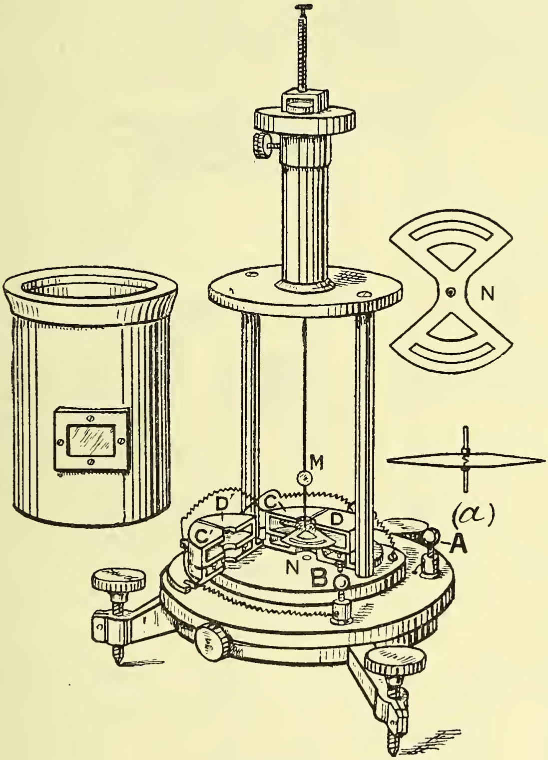


FIG. 268.

insulated and separated from one another by a small air-gap. In the figure, two of the quadrants, C′ and D′, are shown partly removed so as to enable the needle to be seen. The quadrants C, C′ are connected together and to the insulated terminal B, while the quadrants D, D′ are connected to the terminal A. A small mirror M attached to the stem of the needle is used to reflect a beam of light on to a scale, enabling the rotation of the needle to be observed.

If the needle is charged, say positively, and the terminals A and B are

connected to the two conductors of which the difference of potential is to be measured, the needle will be deflected to an amount proportional to the difference of potential. The reason why the needle is deflected is that if, say, A is at the higher potential, the quadrants D and D′ will receive a positive charge and the quadrants C and C′ a negative charge. Hence the positively charged needle will be repelled by the positively charged quadrants D and D′ and attracted by the negatively charged quadrants C and C′. The needle will be deflected till the couple due to the torsion of the suspending wire is equal and opposite to the couple produced by the action of the charged quadrants.

A modified form of quadrant electrometer, called an electrostatic

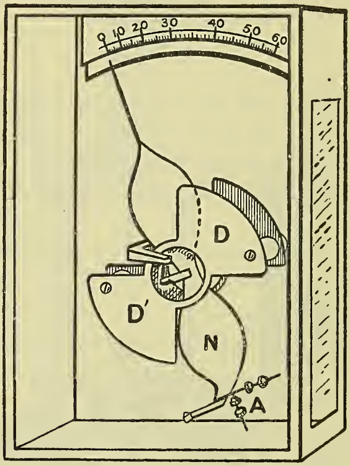


FIG. 269.

voltmeter, is shown in Fig. 269. Here the needle N is supported on a horizontal axle, and there are generally only a single pair of quadrants, D, D′. The centre of gravity of the needle is slightly below the axle, and hence when the needle is deflected there is a turning moment, tending to bring it back to its zero position, due to gravity. By adjusting the weights A the position of the centre of gravity of the needle, and hence the force of restitution when the needle is deflected, can be adjusted. In this way the sensitiveness of the instrument can be varied.

Electrostatic voltmeter.

In some forms of this voltmeter (called multicellular) there are a number of needles mounted on the same axle, each needle being provided with a pair of quadrants. In these voltmeters the *needle* is connected to one of the points between which the potential difference is to be

measured, and the other point is connected to the pair of quadrants. Thus the needle and quadrants receive opposite charges and attract one another.

147. Electrical Machines.—The amount of charge which can be obtained by rubbing a body such as sealing-wax or glass is not great, even when a cylinder or disc is rotated against a rubber of silk, and if any but a very small charge is required a machine the action of which depends on induction must be used.

The Electrophorus.

The simplest form of induction electrical machine is the electrophorus. This instrument is shown in Fig. 270, and consists of a disc of resin or ebonite, AB, and a metal plate, CD, which is attached to an insulating handle, E, by means of which it can be raised from the disc and carried about. The disc is electrified by friction, and the plate is placed on the top. Suppose that the material of the disc is such that it becomes positively electrified on friction, so that when thus electrified we shall have a number of tubes of force stretching from the disc to the walls of the room. On account of the fact that the surface of the disc is never quite plane, when the plate is placed on the top, contact will only take place at a very few points. Thus the plate does not become appreciably electrified by conduction from the disc, for, as it is an insulator, the electrification from those parts which are not in immediate contact with the plate are not able to travel up to the disc. Hence those lines of force which, before the plate was placed over the disc, stretched from the upper surface of the disc to the walls of the room now terminate on the lower surface of the plate, while fresh lines start from its upper surface and stretch away to the walls. If now the plate is earthed, that is, is put in conducting communication with the walls, the lines which start from the upper surface of the plate will be able to shorten and vanish; that is, there will now only be the lines which, starting from the upper surface of the disc, terminate on the lower surface of the plate. If now the plate is lifted up from the disc by its insulating handle, the distribution of the lines of force will alter. Some of the lines will still stretch from the disc to the plate, but as the plate gets further and further from the disc the number of these lines gets less and less. The other lines are, on account of the repulsion they exert on one another, driven out sideways till they meet the walls, their positive ends still remaining on the disc and their negative ends still remaining on the plate. When a line meets the wall it will break in two, and we shall have two separate lines, one starting from the disc

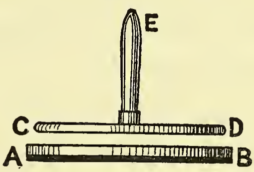


FIG. 270.

and ending on the wall, and the other starting from the wall and ending on the plate. The lines which still stretch from the disc to the plate correspond to that fraction of the charge of the plate which is still "bound" by the charge on the disc, while the lines which stretch from the plate to the walls correspond to the "free" charge of the plate. When the plate is in contact with the disc and has been put to earth its potential is zero. As it is raised up from the plate, having been insulated, its potential will gradually increase; that is, in the case we have supposed, since the charge of the plate is negative, its potential will fall more and more below that of the earth.

If, after having been removed to a distance from the disc, the plate is put to earth, the lines which start on the walls and terminate on the plate will be able to shorten and vanish, and the plate will be discharged. Now in the series of operations we have performed the charge on the disc has not been affected, and hence the plate may be replaced and the whole cycle of operations gone through again, and so on, so that the plate may be charged any number of times without recharging the disc.

It may at first sight seem as if in this way we were able to produce an indefinite amount of electricity without doing any work, and since we have seen that a charged conductor possesses energy in virtue of its charge, this would be contrary to the doctrine of the conservation of energy. It must, however, be remembered that when the plate is in contact with the disc its potential is zero after it has been put in communication with the earth, and it does not then possess any available charge. It is only after the plate has been removed from the vicinity of the inducing charge that it possesses any "free" charge. Now in order to move the plate away from the disc, work has to be done against electrical attraction between the inducing and the induced charges, or, in other words, work has to be done to stretch out the lines of force, and it is this work which is the equivalent of the energy of the electrical charge on the plate. We are therefore here directly converting the mechanical work done by our muscles when we raise the plate into electrical energy.

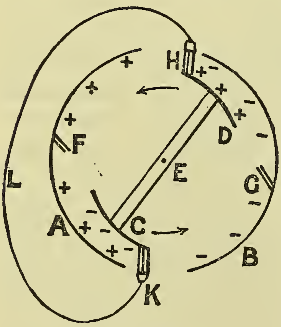


Fig. 271.

In the electrophorus a number of separate operations have to be gone through each time the plate is charged, and it naturally occurs to one to try and invent an arrangement by means of which these operations are performed automatically. The simplest of these is that due to Lord Kelvin, which is known as

Thomson's replenisher. It consists of two metal plates, A and B (Fig. 271), bent so as to form portions of a cylindrical surface, these plates being carried on insulating supports. Two small metallic brushes, F and G, are attached to the inside of A and B. Two other metallic plates, C and D, are carried by an insulating arm, E, which is pivoted so that it can turn about an axis perpendicular to the plane of the figure. Lastly, there are two other metal brushes, H and K, which are in metallic connection with one another by the wire L. The brushes are so arranged that as E rotates, the plates C and D make contact with them.

Suppose that by means of an electrified rod the plate A is given an initial small positive charge, while the plate B is given a small negative charge, and that the movable arm is in the position shown in the figure. Owing to the inductive action of the charged bodies A and B, the plates C and D become electrified, one positively and the other negatively, for they form a single conductor on account of the connecting wire L. Thus we have lines of force starting from D and ending on B. If the arm E is now rotated in the direction of the arrows, the first thing that happens is that, as D and C move round, the connection between them through the brushes H and K and the wire L is broken, while the lines of force are drawn out. The drawn-out lines of force will, on account of their mutual repulsion, spread out, and so most of them will come in contact with the metal plate A. Each line, when it touches A, will divide into two parts, one part stretching from D to A, and the other from A to B, or even by further subdivision from A to the walls. As on this account as many new lines will enter A as leave it, the charge on A will be unaltered. When D touches the brush F it becomes virtually a part of the conductor A, and thus the lines which stretch from D to A contract to nothing; that is, the lines which terminate on A vanish, and so on account of the new positive lines, which were added to A when the lines stretching from D to B split up, the charge on A is increased. In the same way the negative charge on C is transferred to B. As the rotation is continued, the plate D comes into the position in which the plate C is shown in the figure, and the whole process is repeated. Thus by the continuous rotation of the arm E carrying the two plates the charges on the conductors A and B are increased, the one being charged positively and the other negatively.

Although we have supposed an initial charge to be given to A and B, the infinitesimal charge which is induced by the friction of the movable plates on the brushes is generally sufficient to start the machine, this small charge being then increased in the manner described. If the movable arm is rotated in the opposite direction the charges on A and B are decreased, so that the arrangement is used in some instruments for

adjusting the charge to a given value, for, by turning the replenisher in one direction or the other, the charge on a body connected to A or B can be increased or decreased at will.

The large electrical machines (Wimshurst, Hotz, &c.) depend for their action on a cycle of operations similar to that described above, a number of inductors being used in place of the single pair (CD, Fig. 271). A description of the way in which these machines work will be found in Watson's *Text-Book of Physics*, § 471.

148. Atmospheric Electricity.—If to the end of an insulated wire is attached a burning match, or if water is allowed to drop from an insulated vessel attached to the end of the wire, the wire is found to acquire the potential of the air surrounding the match or water jet. The action of the jet is as follows: Each drop of water as it forms at the end of the jet is at the potential of the wire, and if this is different from that of the surrounding air then the drop will be charged by induction. If, say, the potential of the air is higher than that of the wire, a negative charge will be induced on the drop, the corresponding positive charge being on the more remote parts of the wire. When a drop breaks away it carries its negative charge with it, but leaves the positive charge on the wire, and hence the potential of the wire is raised. This action goes on till the potential of the wire is equal to that of the air at the point where the drops separate. If the wire is connected to one pair of quadrants of a quadrant electrometer, the other pair being connected to earth, the deflection of the electrometer will measure the difference in potential between the wire and the earth, or, since the earth is at zero potential, will give the potential of the air at the point where the water jet is placed.

By arrangements similar to that described above the electrical state of the atmosphere has been examined, and it is found that in general, in fine weather, the atmosphere is at a higher potential than the earth, the potential increasing rapidly with height. The potential, however, is very variable, particularly in stormy weather, while when a thunderstorm occurs violent fluctuations occur, the potential sometimes being positive and sometimes negative.

The cause of the atmospheric electricity is not known, though the investigations of radio-activity seem inclined to throw some light on the matter.

The clouds are generally electrified, and if they become charged to a sufficient extent, a spark discharge may take place between two oppositely charged clouds or between a cloud and the earth. This spark discharge constitutes the lightning flash. The functions of a lightning-conductor are first to tend to prevent a cloud in its immediate vicinity acquiring a dangerously high charge, and secondly to furnish a path

along which the discharge may take place if such a discharge occurs. The first of these functions is performed owing to the conductor becoming charged inductively by the cloud and the induced electricity of opposite sign to that of the cloud streaming off from the sharp point. This electricity neutralises in part the charge on the cloud. The end of a lightning-conductor is buried in a damp spot, as damp earth is a good conductor.

CHAPTER V

THE ELECTRIC CURRENT

149. The Electric Current.—If we have two conductors at different potentials, and put them in conducting communication by means of a wire, there will be a redistribution of the electrical charges on the conductors, positive electricity leaving the conductor at the higher potential and increasing on the other till the potentials of the two conductors become equal. If by any means we were able to keep up the difference of potential between the two conductors, although they are connected by the wire, then this transference of electricity would continue and, as we shall see, the wire will possess certain properties which are dependent on the transference of the electricity. In these circumstances the wire is said to be traversed by an electric current. The current is assumed to flow in the direction from the body at the higher potential, through the wire, to the body at the lower potential. The word current was originally used when electricity was regarded as a fluid which flowed from the conductor at the higher potential through the wire, just as a fluid flows from a place at a higher level through a pipe to a place at a lower level. The only thing the passage of which we are able to recognise, however, is energy, this energy being in the form we call electricity, but of the nature of which we are ignorant; and so far from the energy being transmitted by the wire through which the current is flowing, the accepted belief nowadays is that the energy is really transmitted by the insulating dielectric which surrounds the wire, and that the function of the wire is to *direct* the flow of energy. Keeping this warning in mind, it will be permissible to speak of a current of electricity flowing through a wire, and to refer to the phenomena in the space surrounding the wire as due to this current, although we no longer by these terms mean to imply any supposition as to electricity being of the nature of a fluid, or as to the wire being the path along which the energy flows.

The cause of the electric current in the wire is the fact that the two ends of the wire are at different potentials, and in such a case, where the effect of the difference of potential is to produce an electric current, that is, to move positive electrification from one place to another, it is generally spoken of as an electromotive force. Thus what we have hitherto spoken of as the difference in potential between two bodies will

often, when we are dealing with electro-kinetics, be called an electromotive force between the two bodies. It must, however, be remembered that electromotive force and difference of potential[1] are two different names for one and the same thing, and the restriction of the one term more or less rigorously to electrostatics and of the other to electrokinetics is simply a matter of usage.

The electromotive force producing the current might be supplied by connecting the body with one of the terminals of an electrical machine, and under these conditions the electrical machine can be regarded as a source of electromotive force. The detailed study of the other sources of electromotive force can be undertaken with more profit at a later stage, so that for the present it will be sufficient to suppose that what is called an electric battery or voltaic cell is employed. One of the simplest of such cells is that due to Daniell, and consists of a plate of copper immersed in a solution of copper sulphate, and a plate of zinc immersed in a solution of zinc sulphate or in dilute sulphuric acid, the two solutions being separated from one another by a partition of porous earthenware. When the copper plate is connected with the zinc plate by means of a conducting wire, this wire will be traversed by a current. If the copper and zinc plates are connected with the opposite quadrants of a quadrant electrometer the needle will be deflected, and show that the copper is at the higher potential.

The consideration of the manner in which the electromotive force in this cell is developed is postponed, but it may be of use to say that the energy necessary for the maintenance of the electric current in the wire connecting the copper to the zinc is due to the chemical changes which go on in the cell when a current is passing, for the zinc is dissolved forming zinc sulphate, while the copper sulphate solution is decomposed, the copper being deposited. We have seen in § 75 that in every chemical reaction there is a definite quantity of heat absorbed or evolved, and in this case more heat would be evolved in the conversion of a given quantity of zinc into zinc sulphate than is absorbed by the splitting up of the quantity of copper sulphate solution which occurs in the same time, and it is from this surplus energy that the energy necessary for the maintenance of the electric current is derived.

150. Oersted's Experiment. Unit Current and Electromagnetic System of Units.—Hitherto we have not had to deal with any phenomenon connecting magnetism and electricity, although some of the points in which the two classes of phenomena resemble one another may have suggested that some connection must exist. The honour of being the first to discover any connection between electricity

[1] The letters E.M.F. will be employed as an abbreviation for electromotive force and P.D. for potential difference.

and magnetism belongs to Oersted, who found that a conductor in which a current is flowing exerts an action on a neighbouring magnetic needle. If a wire is stretched horizontally in the magnetic meridian, so as to be vertically over a pivoted magnetic needle, then the needle is deflected if a current is passed through the wire, and tends to set itself at right angles to the wire. On reversing the direction in which the current is flowing in the wire, the direction in which the north pole of the needle is deflected is also reversed. The direction of the deflection is also reversed if, instead of being placed over the needle, the wire is placed below the needle.

A number of rules have been given to remember the direction in which the needle is deflected by a conductor carrying a current in a given direction, the two most commonly employed being the following:

1. Imagine yourself swimming in the wire in the direction in which the current is flowing, and facing the magnetic needle; then the north pole will be deflected towards your left hand, the south pole being deflected in the opposite direction (Ampère's rule).

2. Place your right hand alongside the wire, with the fingers pointing in the direction in which the current is flowing, so that the palm of the hand is turned towards the magnet, then the outstretched thumb will point in the direction in which a north pole will be deflected.

Lines of force due to a current flowing in a wire.

If a wire through which a fairly strong current is passed is held in a vertical position, so that it passes through a hole in a horizontal plate of glass, and iron filings are scattered over the glass, on tapping the glass the filings will set themselves in curves which, as in the case of the field of a magnet, indicate the direction of the magnetic lines of force. A series of curves obtained in this way are shown in Fig. 272, and it will be seen that the lines of force consist of a series of circles, the axis of the wire being at the centre of each.

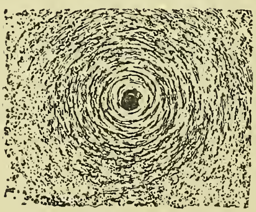


FIG. 272.

The direction in which the lines of force run can be at once obtained from either of the rules as to the direction in which a north pole is deflected in Oersted's experiment. Suppose that the current in the wire of Fig. 272 is running in the direction from beneath the paper to above, then, to a person swimming with the current, a north pole will be deflected to the left hand; thus the lines of force run in the anti-clockwise direction. If we imagine that a corkscrew is placed in the place of the wire conveying the current, with its point in the direction in which the current is flowing, and is then turned in the direction in which it is turned in order

to drive it into a cork, it will travel forward in the direction in which the current is flowing, and the direction in which it is turned will be the direction in which the lines of force of the current run. Thus the direction of the electric current, and the sense in which the lines of force run, are related to one another in the same way as are the direction of motion of an ordinary right-handed corkscrew, or other kind of screw, and the sense in which it is turned. It is of great importance, for following the forces in play between a conductor conveying a current and a magnet, or another conductor which is also conveying a current, to learn to be able at once to tell in which direction the lines of force in the neighbourhood of the conductors are running.

If, instead of being straight, the conductor is bent into the form of a circle, the lines of force all thread through the space enclosed by the conducting hoop.

Since the space in the neighbourhood of a conductor in which a current is flowing is, owing to the current, a magnetic field, and the strength of this magnetic field can be measured by the methods given in the preceding pages, we may take the strength of the field at a given distance from the conductor, which has a given shape, as a measure of the strength of the current flowing in the wire. A system of electrical units has been derived in this way, the starting-point being the strength of the magnetic field due to a conductor conveying the current. The conductor is supposed to be in the form of an arc of a circle of which the radius is one centimetre, the length of the arc being also one centimetre. The unit current is such that the magnetic field produced at the centre of the circle, of which the conductor is an arc, is unity. Hence, since the unit magnetic field is such that the unit north pole is acted upon by a force of a dyne, the unit current may be defined in this system as such that if flowing in the arc of a circle of which the radius is one centimetre, the length of the arc also being one centimetre, then the force exerted on a unit pole placed at the centre of the circle will be a dyne.

Electro-magnetic unit of current.

The unit current thus defined is called the C.G.S. unit of current on the *electro-magnetic system*, and this, or at any rate a multiple of it, is the one almost exclusively employed. Another unit of current can be derived from the unit quantity of electricity as defined in § 140. Such a unit corresponds to the passage of unit quantity of electricity in a second, and is called the electrostatic unit of current.

In the electro-magnetic system the unit of electromotive force, or difference of potential, is such that to cause a unit current to flow between two points on a wire between which unit difference of potential exists involves the performance of one erg of work. Hence if a current

C traverses a wire, and the P.D. between the ends of the wire is E, the work performed is CE.

The units of current and electromotive force given above are of rather inconvenient magnitude, and hence in practice multiples of these are generally employed. These practical units we shall consider later (§ 155).

151. Ohm's Law. Resistance.—If when a current C is passed through a wire and the difference of potential E between the ends of the wire is measured, Ohm found that the ratio E/C is constant for the given wire, so long as its temperature remains constant, whatever the values of the current. This constant ratio, which is a physical property of the given conductor, is called its *resistance*. Calling the resistance R, we have

$$R = E/C \quad . \quad . \quad . \quad . \quad (163)$$

and this is the symbolical representation of *Ohm's law*.

A conductor has unit resistance when unit electromotive force causes unit current to flow through the conductor.

The resistance of a given wire depends on the dimensions of the wire and on the material of which it is composed. It is found that the resistance of a uniform wire is directly proportional to its length l and inversely proportional to its cross-section s, so that

$$R = kl/s \quad . \quad . \quad . \quad . \quad (164)$$

where k is a constant, which depends on the nature of the material of which the wire is composed, and is called the *specific resistance* or *resistivity*[1] of the material. If both l and s are equal to unity $R = k$. Hence k is the resistance of a wire of unit length and unit cross-section.

It is sometimes useful to deal with the reciprocal of the resistance of a conductor, and this quantity is called its conductance. If S is the conductance, so that $S = 1/R$, Ohm's law may be written

$$C = ES \quad . \quad . \quad . \quad . \quad (165)$$

The specific conductivity of a material is equal to the reciprocal of the specific resistance.

If two wires of which the resistances are R_1 and R_2 respectively are placed end to end and a current C is passed through the two, the difference of potential between the ends of the first is R_1C, and that between the ends of the second is R_2C. Hence, since the potential of the end of the first is equal to that of the beginning of the second, for

[1] Names terminating in -ance, such as resist*ance*, are generally used to indicate the property of some particular body, *cf.* the resistance of a given wire, while names ending in -ity are used to indicate the property of a given material, *cf.* the resistiv*ity* of copper.

they are in contact, the difference of potential E between the ends of the combination is given by

$$E = R_1C + R_2C = (R_1 + R_2)C.$$

But if R is the resistance of the combination $E = RC$. Hence it is evident that the resistance of two conductors placed end to end, or in *series* as it is called, is equal to the sum of the individual resistances.

Next suppose the two wires are connected together at each end, as shown at A and E (Fig. 273), and that a current C is passed through the

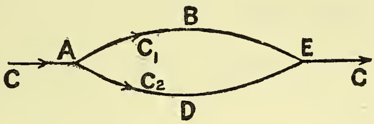


FIG. 273.

Resistance of conductors in parallel.

combination, entering at A and leaving at E. The wires are now said to be arranged in *parallel* or *multiple arc*. Let E be the P.D. between the points A and E, and C_1 the current in the branch B, and C_2 that in the branch D. Then from Ohm's law, $C_1 = E/R_1$ and $C_2 = E/R_2$. Hence, since C_1 and C_2 are together equal to the total current C passing through the combination, we have

$$C = C_1 + C_2 = E\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \quad . \quad . \quad (166)$$

Hence the same current passes through the combination as if they were replaced by a single wire of which the resistance was

$$\frac{1}{1/R_1 + 1/R_2} \text{ or } \frac{R_1R_2}{R_1 + R_2}$$

Since $1/R$ is the conductance, this result may be expressed as follows: The conductance of two wires in parallel is equal to the sum of their separate conductances. If there are any number of conductors in parallel the conductance of the arrangement is equal to the sum of the conductances of the separate conductors.

Returning to the case of the two wires, we see that

$$\frac{C_1}{C_2} = \frac{R_2}{R_1}$$

or adding the denominator of each fraction to the numerator to form a new numerator

$$\frac{C_1 + C_2}{C_2} = \frac{R_1 + R_2}{R_1}$$

or

$$\frac{C}{C_2} = \frac{R_1 + R_2}{R_1}$$

$$\therefore \qquad C_2 = \frac{R_1 C}{R_1 + R_2}$$

Similarly
$$C_1 = \frac{R_2 C}{R_1 + R_2}$$

Thus by measuring the current in one branch we can calculate the total current which is passing through the system. This result is utilised when what is called a *shunt* is employed with a current measuring instrument. Thus if the current to be measured is too great for the instrument a resistance is placed in parallel with the instrument so that only part of the current actually passes through the instrument. If G is the resistance of the measuring instrument, S that of the shunt, and c is the value of the current as read off, the total current C is given by

Shunts.

$$C = \frac{S + G}{S} c \qquad . \quad . \quad . \quad . \quad (167)$$

If S is made equal to $G/9$, then

$$C = 10c,$$

or the actual current is ten times the reading. In the same way by using a shunt of resistance $G/99$, the current is 100 times the reading, and so on.

152. The Wheatstone's Bridge.—A system of conductor's, AE, EC, AD, DC, arranged as in Fig. 274, the points A and C being connected with the poles of a battery, B, and the points D and E being connected through an instrument for showing whether a current is passing, called a galvanometer, G, is known as a Wheatstone's network of conductors, or Wheatstone's bridge. If the resistances of the separate conductors are as shown on the figure, and these resistances are so adjusted that no current passes through the galvanometer, then the following holds :—

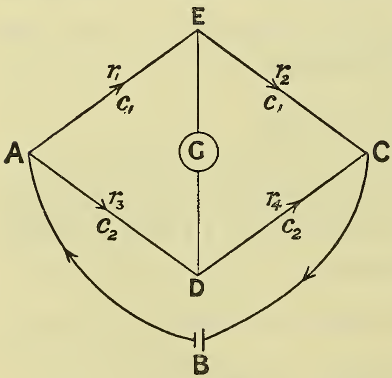


FIG. 274.

$$r_1/r_2 = r_3/r_4.$$

Since there is no current through the galvanometer, the potential of the points E and D must be the same. Further, since the ends of the resistances at A are connected together, and therefore must be at the same potential, and similarly at C, it follows that the difference of potential, e_1, between A and E is equal to that between A and D, and the difference of potential, e_2, between E and C is equal to that between

D and C. Since the current, c_1, passing through AC is the same as that passing through EC, we have $c_1 = e_1/r_1 = e_2/r_2$,

or $$\frac{r_1}{r_2} = \frac{e_1}{e_2}$$

Similarly $$\frac{r_3}{r_4} = \frac{e_1}{e_2}$$

Hence $$\frac{r_1}{r_2} = \frac{r_3}{r_4} \quad . \quad . \quad . \quad . \quad . \quad (168)$$

This result may be written $r_1 = r_2 \cdot \frac{r_3}{r_4}$, which shows that if we know the value of the resistance r_2, and also the *ratio* of the resistances r_3 and r_4, we can calculate the value of the resistance r_1. Thus if we know the value of one resistance and the ratio of two others, which, when arranged together with an unknown resistance so as to form a Wheatstone's net, give no current in the galvanometer, we can immediately calculate the value of the unknown resistance.

If the battery is connected between E and D and the galvanometer

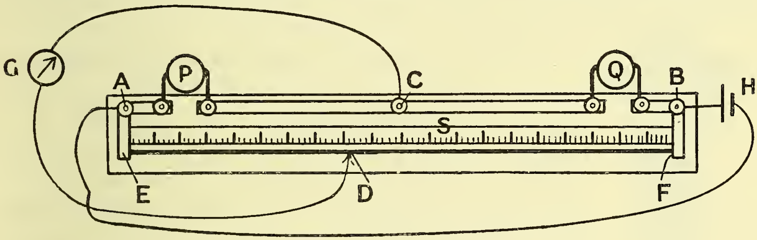


FIG. 275.

between A and C, it can be shown by an argument similar to that given above, that when the galvanometer is undeflected the resistances must satisfy the same condition, viz. $r_1/r_2 = r_3/r_4$.

A simple form of Wheatstone's bridge is shown in Fig. 275. It consists of a wire EF, generally a metre long, stretched alongside a divided scale S and attached to two thick copper strips A and B. The unknown resistance P and the known resistance Q are connected to terminals attached to these strips and a third strip C. The battery is connected to A and B and one terminal of the galvanometer to C. The other terminal of the galvanometer is connected to a contact key D which can be moved along the wire. The position of D is adjusted till on making contact with the wire no deflection of the galvanometer is produced. Since the wire EF is of uniform cross-section, the resistance of the part ED is to the resistance of the part DF as the *length* ED is to the *length* DF. Hence $$\frac{P}{Q} = \frac{ED}{DF}.$$

The lengths ED and DF are read off on the scale, and hence if we know Q we can at once calculate the value of P.

Another form of Wheatstone's bridge, known as the Box or Post-Office form of bridge, has no stretched wire. In this form the ratio of the resistances r_3 and r_4 is not capable of being given any value we please, but the bridge is supplied with a number of coils of wire by means of which certain fixed ratios can be obtained, the usual ratios being 1 : 1, 1 : 10, 1 : 100, 1 : 1000, 1000 : 1, 100 : 1, 10 : 1. In addition to these ratio coils, there are a set of coils by means of which the resistance in the arm EC can be made any whole number of ohms between 1 and 10,000. If the ratio of the proportional arms is 1 : 1, then the resistance unplugged in the third arm will be equal to the resistance being measured. If the ratio of the proportional arms is 1 : 10, then the resistance being measured is ten times the resistance unplugged in the third arm; while if the ratio is 10 : 1, then the resistance is one-tenth, and so on.

153. The Platinum Thermometer.—For measuring temperatures much above 300° the mercury thermometer is quite unsuited, and although the air thermometer can be employed, yet its use is accompanied by so many experimental difficulties as to render it only suited for standardising other more handy forms of thermometer. The fact that the resistance of a pure metal changes to quite a considerable extent with change of temperature, and that the resistance of a conductor can be measured with a Wheatstone bridge with great accuracy and ease, has been utilised as a means of measuring temperature. Callendar first showed that if care is taken to protect a platinum wire from the action of certain gases, the change of resistance with temperature is very constant, and hence may be used for measuring temperature.

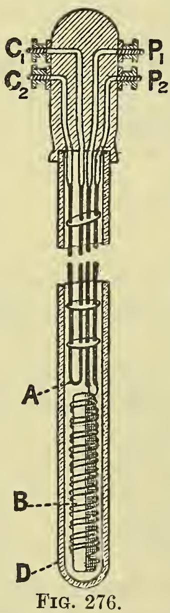


FIG. 276.

The form of platinum thermometer which he devised is shown in Fig. 276. It consists of a wire of pure platinum wound on a thin mica frame and enclosed in a glass, or, if it is required for measuring high temperatures, in a porcelain tube. The ends of the wire are connected to two thick platinum leads, P_1, P_2, by welding the platinum together. Flexible copper wires are used to connect the platinum leads to a Wheatstone's bridge, by means of which the resistance of the wire can be measured.

Since any change of temperature would affect the resistance of the platinum and copper leads, while what we require to measure is the

change of resistance of the coil of thin wire only, Callendar has introduced a compensating device. This consists of a second pair of leads, C_1, C_2, of exactly the same resistance as the others, but which are connected together at A. These dummy leads are connected, by means of a pair of flexible copper leads of the same resistance as the others, with the adjacent arm[1] of the Wheatstone's bridge to that in which the platinum thermometer is placed. Hence, as both sets of leads are placed close together, their temperature will always be the same, and so any change in resistance produced by a variation of the temperature of the room will affect both equally. But an equal small increase in the resistance in the adjacent arms of the bridge will not affect the galvanometer, and so the arrangement will be independent of any change in the temperature of the leads.

It has been found that a platinum thermometer gives consistent results up to a temperature of about 1000° C., and it has been used with much success to measure the melting-point of metals.

If the resistance of a platinum thermometer at 0° C. is R_o and that at 100° C. is R_1, the quantity $R_1 - R_o$ is called the fundamental interval for the thermometer. If the resistance at any temperature t is R_t, then the quantity p_t obtained from the relation

$$p_t = 100\frac{R_t - R_o}{R_1 - R_o} \quad . \quad . \quad . \quad . \quad (169)$$

is called the platinum temperature corresponding to the temperature t. If the change in resistance of platinum were exactly proportional to the change in temperature, then the platinum temperature would be the same as the temperature on the ordinary air scale. This, however, is not the case, and Callendar has shown that the platinum temperature p_t is connected with the temperature t on the air scale by the relation

$$t - p_t = D\left\{\frac{t}{100} - 1\right\}\frac{t}{100} \quad . \quad . \quad . \quad (170)$$

where D is a constant for any particular wire. By this relation temperatures on the platinum scale can be reduced to the air scale.

154. Joule's Law.—When a current is passed through a conductor heat is developed, and Joule, who first investigated this effect, showed that for a given conductor the heat is proportional to the square of the current, and this is known as Joule's law.

Joule's law at once follows from Ohm's law, and the relations between

[1] When these compensating leads are employed the ratio arms of the bridge must be of equal resistance.

the units of current and P.D. we have given on p. 443. Thus if the resistance of a conductor is R and a current C is passing, the P.D. E between the ends of the conductor is given by $E=RC$. Now when a current C flows for one second, C units of electricity will pass through the conductor, and as by definition E ergs of work is performed when the unit quantity of electricity passes between two points between which there exists a difference of potential E, it follows that EC ergs of work are required to send the current C through the wire. Now this work appears as heat developed in the conductor, so that the heat developed per second is EC, or RC^2 since $E=RC$. Thus if H is the heat, measured in ergs (see p. 181), developed in a time t, we have

$$H=C^2Rt \quad . \quad . \quad . \quad . \quad (171)$$

The quantity of energy which becomes converted into heat when a given current flows through a given conductor is independent of the direction in which the current flows, for, as the current always flows from the point at the higher potential to the point at the lower potential, if we reverse the direction in which the current is flowing, this means that we have reversed the direction in which the P.D. is acting; and since the resistance of a conductor is independent of the direction in which the current is flowing, the conditions, as far as the work done by the current and hence the heat developed, are exactly the same when the current is reversed as they were before. Thus the passage of a current through a conductor of finite resistance is always accompanied by the conversion of a definite quantity of electrical energy into the form of heat. Since when the current is reversed the conversion into heat continues at the same rate as before, this conversion of electrical energy into heat, when a current passes through a conductor, is an irreversible process (§ 82). As we shall see later, there are conditions under which heat developed at a given point due to the passage of a current is a reversible process, so that on reversing the current heat is now absorbed at the point; in the case of heat developed according to Joule's law, however, this is never the case. In many cases the heat produced according to Joule's law is simply a waste of energy, thus it is important to reduce it to a minimum. This can be done, if we suppose that a given current has to be transmitted, by reducing the resistance of the conducting wires.

If J is the mechanical equivalent of heat, the heat H^1 developed by a current C in a resistance R in a time t is given in calories by the relation

$$H^1=\frac{C^2Rt}{J} \quad . \quad . \quad . \quad . \quad (172)$$

Hence if H^1, C, and R are measured, the value of the mechanical equivalent can be calculated. This is the principle of a method of measuring J used by Griffiths. His apparatus consisted of a coil of platinum wire through which the current could be passed, and which had two wires attached, so that the difference of potential between the ends of the coil could be measured. This coil was contained inside a closed calorimeter, which was itself placed inside a large steel chamber, the space between the outside of the calorimeter and the walls of this vessel being exhausted of air so as to reduce the loss of heat due to convection. The calorimeter contained, in addition to the coil, a stirrer, which was rotated at a high speed, so as to ensure the water inside being thoroughly well mixed. The temperature of the water in the calorimeter was measured with a platinum thermometer, and the resistance of the coil at different temperatures was determined so that, knowing the P.D. between the terminals and the temperature, the resistance of the coil and the rate at which heat was being developed by the current could be calculated. A certain amount of heat was also developed by the friction of the stirrer against the water. The amount of heat thus developed at different rates of stirring was determined by making observations of the rise in temperature of the calorimeter, due to the stirring alone, when no current was passing through the coil. The water value of the calorimeter and of the stirrer and coil was determined by making experiments with various quantities of water in the calorimeter.

Determination of the value of J from the heat developed by a current.

The heat developed in a conductor is made use of in the fuses included in electric lighting circuits; the function of which is to break the circuit if an excessively large current passes. These fuses consist of a wire of some fusible metal, such as lead, the diameter being so chosen that when the maximum current which the circuit is capable of safely carrying is exceeded the temperature of the wire reaches the melting-point.

The incandescent filament electric lamp depends for its action on the heat developed by the passage of the current. The filament is enclosed in an evacuated glass bulb to protect the filament against oxidation by the air at the high temperature to which it is raised. In the older type of lamp the filament was of carbon. The modern high-efficiency lamps have metallic filaments, the metals generally employed being tungsten or tantalum. These metals allow of the filament being raised to a higher temperature without the filament being destroyed than is the case with carbon, and hence a greater proportion of the energy supplied to the lamp is given out in the form of *visible* radiation (§ 125).

Incandescent lamps.

When two rods of carbon which are at a potential difference of about

60 volts are momentarily brought into contact, so that a current passes, and then separated, the current continues to cross the space between the rods and a very bright light is emitted. This arrangement constitutes an electric arc. The carbons are gradually consumed owing to oxidation by the air, the carbon which forms the positive pole being eaten away the faster. If an image of the arc is projected on a screen, or if the arc is examined through a dark glass, the end of the positive carbon will be seen to be slightly hollow, and most of the light is emitted from this hollow, which is called the crater of the arc. With plane carbons only a very small proportion of the light comes from the arc which occupies the space between the carbons. If, however, a core of salts of certain salts, chiefly calcium fluoride, is placed inside the carbon rod a very considerable amount of light is emitted by the arc itself. In these flame arcs, as they are called, the colour of the light depends on the nature of the metals present, and if the arc is examined with a spectroscope the characteristic bright lines of the metals present will be observed.

The electric arc.

Another practical application of the heat developed by the passage of a current is the electrical furnace, by means of which higher temperatures can be attained than is possible with the ordinary type of furnace.

155. The Practical System of Electrical Units.—In the C.G.S. system of electro-magnetic units the values of some of the units are inconveniently small. Thus the potential difference at which electricity is supplied in the majority of electric lighting systems is 200000000 C.G.S. units. For this reason a system of multiples of the C.G.S. units has been adopted, called the practical system of electro-magnetic units.

The values of these practical units in terms of the C.G.S. units and the names they have received are given in the following table:—

Quantity.	Name of Practical Unit.	Equivalent in C.G.S. Units.
Current	Ampere . .	10^{-1} electro-mag. units.
Quantity	Coulomb . .	10^{-1} ,, ,,
Electromotive force . .	Volt . . .	10^{8} ,, ,,
Resistance	Ohm . . .	10^{9} ,, ,,
Capacity	Farad . . .	10^{-9} ,, ,,
Inductance (self or mutual) .	Henry . . .	10^{9} ,, ,,
Energy or work . . .	Joule . . .	10^{7} ergs.
Power	Watt . . .	10^{7} ergs per second.

It will be observed that the relations between the units themselves in this system are the same as those which exist in the C.G.S. system. Thus take the case of Ohm's law, a current of 10^{-1} C.G.S. will pass when

an E.M.F. of 10^8 C.G.S. acts through a resistance of 10^9 C.G.S., for $10^{-1}=10^8/10^9$. That is, if A is the current in amperes, R the resistance in ohms, and V the E.M.F. in volts, we have $C=E/V$.

Similarly, a current of A amperes passing through a resistance of R ohms will develop A^2R joules of heat per second, that is, the power expanded will be A^2R watts.

When considering the thermal effects of currents it is sometimes convenient to express the heat in terms of the ordinary thermal units. Since one joule is equal to 10^7 ergs, and one calorie is equal to $4{\cdot}186 \times 10^7$ ergs, one joule is equal to $10^7/4{\cdot}186 \times 10^7$ or 0·2389 calories. A joule is equal to ·000948 B.Th.U.

For some purposes even the units of the practical system are of inconvenient size, and hence other units are employed. Some of these are a million times as great and others a millionth of the practical units. These units are indicated by the prefixes *mega-* and *micro-* respectively. Units a thousand times as great or a thousandth of the practical units are indicated by the prefixes *kilo-* and *milli-* respectively. The chief of these auxiliary units are given in the following table :—

Auxiliary Units in the Practical System.

Name.	Equivalent in the Practical System.	Equivalent in C.G.S. Units.
Megohm . .	10^6 ohms	10^{15} electro-mag. units.
Microfarad .	10^{-6} farad	10^{-15} ,, ,,
Microvolt . .	10^{-6} volt	10^2 ,, ,,
Microampere .	10^{-6} ampere	10^{-7} ,, ,,
Kilowatt . .	10^3 watts	10^{10} ergs per second.
Millivolt . .	10^{-3} volts	10^5 electro-mag. units.
Milliampere .	10^{-3} ampere	10^{-4} ,, ,,

The kilowatt is equal to 1·341 horse-power.

The unit generally employed when considering the supply of electrical energy for lighting or power is the kilowatt-hour. This unit is in England called the Board of Trade unit, and is equal to $3{\cdot}6 \times 10^6$ joules.

Quantities of electricity are sometimes expressed in ampere-hours, *i.e.* the quantity of electricity which passes when a current of an ampere flows for an hour. One ampere-hour is equal to a coulomb divided by 3600.

CHAPTER VI

THE MAGNETIC FIELD PRODUCED BY CURRENTS

156. Magnetic Field due to a Circular Conductor in which a Current is Flowing.—We have defined our unit current in the C.G.S. system as such that if this unit current is flowing in a conductor one centimetre long, bent into an arc of a circle of one centimetre radius, the field at the centre of the circle is one gauss, and we have now to consider how the field varies where the shape and dimensions of the conductor are varied.

We may regard the field at any point in the neighbourhood of a conductor in which a current is flowing as due to the combined action of all the small elements into which the conducting wire may be supposed to be broken up. The effect of any such small element cannot be measured experimentally, since it is impossible to obtain such an element, for the current must be conducted to and away from the element, and the magnetic effect of these conductors would have to be taken into account. Ampère, however, made a long series of experiments on the magnetic field of conductors of different forms, and he deduced from his results what would be the magnetic field of a small element of a wire in which a current C is flowing. If e is the length of the element, and if the direction of the length of the element makes an angle θ with the line joining the centre of the element to the point where the strength of the field is to be measured, and the distance of this point from the centre of the element is r, then the strength of the field is $Ce . \sin \theta/r^2$.

Although the correctness of this expression cannot be directly tested by experiment, yet by its means the strength of the field due to conductors of certain fixed forms has been calculated, and the calculated result has been found to agree with the value obtained experimentally.

As an application of the law, we may employ it to obtain the strength of the field near a wire which is bent in the form of a circle of radius R. First let the point at which the force is to be calculated be the centre of the circle, then the angle between the element and the line joining the element to the centre of the circle is always 90° ; hence $\sin \theta$ is 1 for all the elements. The distance of the point where the strength of the field is to be measured from the elements is also constant, being R the radius of the circle. Hence the strength of the field due to each element of

length e is Ce/R^2. Further, since the direction of the lines of force due to each element is at right angles to the length of the element, the directions of the lines of force at the centre of the circle are all parallel to the axis of the circle. Thus the strength of the field due to the combined effect of all the elements is obtained by simply adding the strength due to each of them separately. The factor C/R^2 being common to all the elements, we have simply to add together the lengths of the different elements of the wire and then multiply the sum by C/R^2. But the sum of the lengths of the elements is the circumference of the circle, that is, $2\pi R$. Hence the strength of the field at the centre of the circle is $2\pi C/R$. If, instead of being a complete circle, the wire only occupies an arc of which the length is equal to the radius R of the circle, the strength of the field at the centre is C/R. If, further, the radius of the circle is one centimetre, the length of the wire also being one centimetre, the strength of the field at the centre is C. Hence if the strength of the field at the centre is unity, the current in the wire is also unity, and this result agrees with our definition of the unit current.

If instead of the wire only forming a single turn there are n turns, all of radius R, and a current C is sent through them all, since the field due to all the turns will be parallel, and the strength at the centre due to one turn is $2\pi C/R$, the strength of the field due to the n turns will be $2\pi nC/R$. If, instead of being measured in C.G.S. units, the current is measured in *amperes*, the strength of the field produced at the centre by a current of A amperes, when flowing in a circular coil of radius R and having n turns, is $\pi nA/5R$.

It will be observed that since the strength of the field is proportional to the product of the number of turns into the current in amperes flowing through the turns, it is immaterial, as far as the strength of the field is concerned, whether we use a large current and few turns, or a small current and many turns. Hence, when considering the magnetic field produced by coils of wire, the term *ampere-turns*, which is equal to the product of the current strength in amperes into the number of turns, is often used.

To obtain the strength of the field at any point on the axis of the circle, other than the centre, we may proceed as follows. Let A and B (Fig. 277) represent the cross-section of the circular conductor by a plane (that of the paper) containing the axis of the circle. Let the distance of the point P from the plane of the circle be d, and the angle made by a line joining P to any point on the circumference of the circle with the axis be θ. Consider an element of the wire of length e at A; the strength of the field due to this element at P will be $Ce/\overline{\text{AP}}^2$, since the line AP is at right angles to the element. Also $\overline{\text{AP}}^2 = R^2 + d^2$, so that the force due to the element is $Ce/(R^2 + d^2)$. Since the lines of force of

the element are circles in the plane of the paper with A as centre, the direction of the force at P is tangential to the circle in the plane passing through P perpendicular to the element described about A as centre and with AP as radius, that is, it is along PD, where PD is at right angles to AP. This force may be resolved into two components, one along the axis of the circle and the other along the line PF at right angles to the axis. Since PD is at right angles to AP, and PF is at right angles to CP, the angle FPD is equal to θ. Hence the component of F along the axis is $\frac{Ce}{R^2+d^2}\sin\theta$, and the component at right angles to the axis is $\frac{Ce}{R^2+d^2}\cos\theta$. If we proceed in the same way for the element of length e at B, the component along the axis will also be $\frac{Ce}{R^2+d^2}\sin\theta$, and the component at right angles to the axis will be $\frac{Ce}{R^2+d^2}\cos\theta$, but in the opposite direction to the component due to the element at A. Thus the

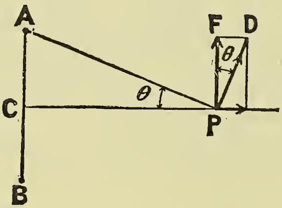


FIG. 277.

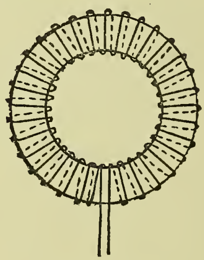


FIG. 278.

components, at right angles to the axis, of the fields due to these two elements are equal and opposite, and hence are in equilibrium and neutralise each other. Since the whole circle may be split up into pairs of elements which bear to one another the same relation as do the elements at A and B, the components at right angles to the axis will on the whole neutralise one another, and in calculating the strength of the field at P we have only to consider the components parallel to the axis. The term $C\sin\theta/(R^2+d^2)$ is common to all the axial components of all the elements, and the sum of the lengths of the elements is the circumference of the circle, that is, $2\pi R$. Hence the strength of the field at P is $2\pi RC\sin\theta/(R^2+d^2)$, or, since $\sin\theta = \text{AC}/\text{AP} = R/\sqrt{(R^2+d^2)}$, this may be written $2\pi R^2C/(R^2+d^2)^{\frac{3}{2}}$. If there are n turns, the strength of the field at P is $2\pi nR^2C/(R^2+d^2)^{\frac{3}{2}}$.

It can be shown that it follows from the definition of the unit current and from Ampère's expression for the field due to an element of current that the magnetic field at a distance r from a long, straight conductor in

which a current C is flowing is equal to $2C/r$. It is to be noticed that the conductor is supposed to be very long compared to r, and that the wires which convey the current to and from the ends of the straight conductor are at a great distance from the point at which the field is measured.

The direction of the field is tangential to the circle drawn through the point in a plane perpendicular to the conductor and with its centre on the conductor. If a unit pole is placed on the circumference of this circle the force acting on it will be $2C/r$, and hence the work which will be performed when this pole is carried round the circle will be $2\pi r \times 2C/r$ or $4\pi C$, and will be positive or negative according to the direction in which the pole is carried round the circle. Since the expression for the work does not involve the radius of the circle, it follows that whatever the radius the work is the same. The reader will easily be able to show from this result that whatever the shape of the path along which the pole is carried, the work performed when it is taken once completely round a closed path which embraces a conductor in which a current C is flowing is $4\pi C$.

Work performed when a unit pole is carried once round a conductor conveying a current.

If there are n conductors in each of which the current C is flowing, say a coil of n turns, then the work performed when the unit pole is carried once round a circuit which threads through the coil is $4\pi Cn$.

The work performed when a unit pole is carried from a point A to another point B is called the *magnetomotive force* along the path AB. If this path is a *closed* curve which threads through the coil considered above, the magnetomotive force is $4\pi Cn$. If the current is A amperes the magnetomotive force is $0{\cdot}4\pi An$, or $0{\cdot}4\pi$ times the ampere-turns.

Magnetomotive force.

A ring (Fig. 278) wound uniformly with insulated wire forms what is called an anchor-ring. If the core is formed of some non-magnetic material (permeability = 1), and has an axial length l, the work performed when a unit pole is taken once round the axis is $4\pi NC$, where C is the current flowing in the wire. If H is the magnetic field along the axis, which from symmetry must be the same all round, the work is Hl. Hence

Field inside an anchor-ring.

$$H = 4\pi NC/l \quad . \quad . \quad . \quad . \quad (173)$$

If n is the number of turns per centimetre of the axis, this result may be written $H = 4\pi nC$.

If we consider a circle in the core parallel to the axis but nearer the centre of the ring of which the length is l', the field strength at all points along this circle is given by $H' = 4\pi NC/l'$. Since l' is less than l, it follows that H' is greater than H, so that the field is not of uniform

strength all over the cross-section of the ring. If, however, the diameter of the ring is great compared to its cross-section, the field strength will be practically uniform throughout the core.

If we imagine the diameter of the ring increased without limit, but that the number of turns of wire per centimetre is kept constant and equal to n, the field inside will always be equal to $4\pi nC$. Now a finite length of such an infinitely large anchor-ring will be straight, and if the length of this straight portion is great compared to the cross-section, the field near the middle will be unaltered if we suppose the remainder of the anchor-ring removed. Hence we conclude that the field near the middle of a straight coil, uniformly wound with n turns per unit length, when traversed by a current C is uniform and equal to $4\pi nC$. It must be remembered that this expression does *not* hold near the ends of the coil. Such a coil is generally called a solenoid.

Field inside a solenoid.

If an anchor-ring has a core of iron of which the permeability is μ, since the magnetising force is $4\pi nC$, the induction through the iron is given by (p. 401)

$$B = 4\pi CN\mu/l \quad . \quad . \quad . \quad . \quad (174)$$

The total number G of lines of force, *i.e.* the flux, which pass round the ring if the cross-section is s is given by

$$G = 4\pi NC\mu s/l \quad . \quad . \quad . \quad . \quad (175)$$

Now $4\pi NC$ is the magnetomotive force round the ring when no iron is present. Hence, calling this quantity M,

$$G = M\mu s/l \quad . \quad . \quad . \quad . \quad (176)$$

Now if an E.M.F. E acts between the ends of a uniform conductor of which the length is l, the cross-section s, and specific resistance k, the current is given by

$$C = Eks/l \quad . \quad . \quad . \quad . \quad (177)$$

Remembering that the E.M.F. is measured by the work performed when unit quantity of electricity flows from one end of the conductor to the other, it will be seen that there is a certain resemblance between equations 176 and 177, and it is on account of this resemblance that the quantity M has been called the magnetomotive force. Similarly, since l/ks is the resistance of the conductor, the quantity $l/\mu s$ has been called the *magnetic resistance* or *reluctance*, so that the total induction, *i.e.* the total number of lines of

The magnetic circuit.

force through a magnetic circuit,[1] is equal to the magnetomotive force divided by the reluctance.

Although this manner of viewing magnetic problems is of considerable use, particularly when dealing with practical problems, such as the design of dynamos and transformers, and has proved suggestive in indicating new paths for experimental research, it must be remembered that the whole analogy is a mathematical one, built up on the similarity of the two equations considered above, there being no physical analogy between the two cases. Thus there is no known magnetic phenomenon which is physically analogous to the conduction current, while physically the analogue of permeability is not specific conductivity but specific inductive capacity. Again, while the resistance of a conductor is independent of the strength of the current, the reluctance depends on the magnetic induction, for, as we have seen in § 138, the permeability of iron varies enormously with the induction.

In the case of the annulus considered above, the tubes of induction are confined to the iron, and the magnetic circuit therefore consists of one medium only. We may, however, apply the idea of the magnetic circuit to cases where the tubes of induction pass through media of different permeability.

In the first place, let us take the case of the iron ring already considered, but suppose that the magnetising coil, instead of being wound uniformly all round the ring, is confined to a small portion of the circumference. In these circumstances, some of the lines of force will leave the iron in the part of the ring which is not covered by the magnetising coil and will travel through the air. Since, however, the permeability of soft iron is several hundred times greater than that of the air, at any rate when the magnetising field is not very great, such a large proportion of the lines will continue all the way through the iron ring, that we may, without making any appreciable error, neglect the ones that do not. If the iron were removed, and a unit pole were carried once round the space previously occupied by the iron ring, it would pass once round each of the turns of the magnetising coil. Hence if there are N turns in this coil, and the current is C, the work done is $4\pi NC$, and therefore the magnetomotive force is $4\pi NC$. Also the reluctance of the iron ring is $l/s\mu$. Thus the flux through the iron is given by

$$G = \frac{\textit{Magnetomotive force}}{\textit{Reluctance}} = 4\pi NCs\mu/l.$$

[1] We have already seen that every line of magnetic force is endless; thus the portion of space through which any set of lines pass forms a closed circuit, and it is therefore known as a magnetic circuit. A magnetic circuit may be formed by one or more different media, and may be single or branched, just as an electrical circuit may be formed by different substances, and may have branches forming loops.

This result is slightly greater than that in the iron which is furthest from the magnetising coil, on account of the lines which thread through the coil, but instead of continuing through the iron, escape into the surrounding air. Still the result obtained is a very near approximation to the truth. The advantage of the magnetic-circuit point of view is apparent if we consider how very difficult it would be to calculate the value of the magnetising force at each point of the iron ring, in order to deduce the induction. The analogue of this problem in electricity would be the case of a ring of copper immersed in a feebly conducting medium, such as water, for in such a case most of the current would traverse the copper, but some would traverse the water, and so the resistance of the circuit would be somewhat less than the resistance of the copper alone, although a very near approximation to the current would be obtained if we neglected the portion of the current which flows through the water.

As another example of the utility of the idea of the magnetic circuit, we may take the case of the iron ring which is lapped over with a uniformly wound magnetising coil, but which at one place has been cut so that the continuity of the iron is broken by a narrow air-gap. As before, the magnetomotive force will be $4\pi NC$. The magnetic circuit is no longer confined to a single medium, but at the gap passes from iron to air. Hence in calculating the reluctance, we have to consider the two portions of the circuit, in one of which the permeability is μ, and in the other portion it is unity. If x is the width of the gap, the length of the iron circuit is $l-x$. Hence the reluctance of the iron part of the circuit is $(l-x)/\mu s$. If the air-gap is at all wide, the lines of force will spread out at the gap, and hence the cross-section of the magnetic circuit in the gap will be greater than s. If, however, the gap is very narrow, the spreading of the lines will be inappreciable, and we may take the cross-section of the circuit at the gap as equal to s. The length of the air part of the circuit being x, the reluctance is x/s. Hence the reluctance of the combined iron and air circuit is

$$(l-x)/\mu s + x/s.$$

Thus the flux through the circuit is given by

$$\begin{aligned} G &= 4\pi NC/\{(l-x)/\mu s + x/s\} \\ &= 4\pi NC/\left\{\frac{l+x(\mu-1)}{\mu s}\right\} \\ &= 4\pi NC\mu s/\{l+x(\mu-1)\}. \end{aligned}$$

If the length of the ring had been $l+x(\mu-1)$, and supposing no gap were present, the flux would have been

$$4\pi NC\mu s/\{l+x(\mu-1)\}.$$

Hence the effect of a gap of length x in reducing the flux is the same as would be produced by a length $x(\mu-1)$ of iron. If the magnetising field $(4\pi NC)$ is 5 C.G.S. units, the permeability for the soft iron, for which the curve in Fig. 253 is drawn, is 2400. Hence if the length of the ring is 30 cm., and its cross-section 4 sq. cm., the flux with a gap a millimetre wide is

$$\frac{5\times 2400\times 4}{30+0{\cdot}1(2400-1)}=\frac{48000}{269{\cdot}9}=177{\cdot}8 \text{ lines.}$$

If no gap were present the flux would be

$$\frac{5\times 2400\times 4}{30}=1600 \text{ lines.}$$

It will thus be seen how enormously the presence of the air-gap reduces the flux.

A bar of iron on which is wound one or more coils of insulated wire becomes magnetised if a current is passed through the wire, and constitutes what is called an electromagnet. By this means we can obtain magnets which are enormously more powerful than any permanent magnets.

157. Galvanometers.—The deflection of a magnetic needle from its position of equilibrium in a magnetic field, either that of the earth or that due to the combined field of the earth and a magnet, by the action of the field due to a wire in which a current is flowing is the commonest way of detecting and of measuring a current. An instrument consisting essentially of a magnetic needle and a conducting wire, so arranged that when a current flows in the wire the needle is deflected and used for detecting or measuring an electric current, is called a galvanometer. Galvanometers may be divided into two great classes, namely, those used for simply detecting the passage of a current or of *comparing* the magnitude of two currents, and those in which, from the magnitude of the deflection, we can calculate the magnitude of the current. In the first class the chief requisite is sensitiveness, that is, that a very small current shall produce a measurable deflection of the needle; while in the second class sensitiveness has to be subordinate to the requirement that we must be able to calculate from the dimensions of the coil, &c., the value of the field produced at the centre by unit current.

As we have seen in the last section, the strength of the field at the centre of a circular coil varies inversely as the radius of the coil, and hence if we wish to make the field produced by a given current as great as possible, we must make the radius of the coil small. Further, since the strength of the field is directly proportional to the number of turns, the sensitiveness will obviously increase as the number of turns increases.

Suppose that the plane of the galvanometer coil is made to coincide with the magnetic meridian, so that the field due to the coil is at right angles to the magnetic meridian, then the direction in which a magnetic needle suspended at the centre of the coil will set itself will be the direction of the resultant of the field of the coil and that of the earth. If, then, by means of an auxiliary magnet the strength of the earth's field, in which the needle hangs when no current is passing, is decreased, the resultant of this field and that due to the current in the coil will be nearer to the direction of the field due to the coil. The deflection of the needle will therefore be greater for a given current in the coil.

The sensitiveness of the galvanometer is also sometimes increased by employing what is called an astatic system for the needle. An astatic system consists of two magnetic needles of almost the same magnetic moment, fixed to a stem to which is attached the fibre by which they are suspended in such a way that their poles are turned in opposite directions, as shown in Fig. 279. If the magnetic moment of the magnet ns is m, while that of $n's'$ is $m+x$, where x is a small quantity, the system will set itself in the magnetic meridian with the pole n' towards the north, for the magnet $n's'$ is the stronger. If the system is deflected through an angle θ from the meridian, the couple tending to bring the needle $n's'$ back into the meridian is (§ 133) $(m+x)H \sin \theta$, while the couple tending to turn the magnet ns out of the meridian is $mH \sin \theta$. Hence the resultant couple tending to bring the system into the meridian is $xH \sin \theta$.

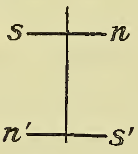


FIG. 279.

By making the quantity x small, this couple can be made as small as we like, so that if the needles are of almost the same magnetic moment, the directive couple acting on the astatic system, due to the field in which the system is suspended, is very small.

In the application of an astatic system to the galvanometer, the coil of wire is either made to surround one needle only, or two coils are employed, one round each needle, but the current is sent round the two coils in opposite directions, so that the field due to the coils in each case tends to deflect the needles in the same direction. Thus while the deflecting couple due to the field of the galvanometer coils remains the same, the directing couple which tends to bring the needles into the meridian, and hence opposes the deflection of the needle, is reduced, and the deflection produced by a given current in the coils of the galvanometer is increased.

In sensitive galvanometers the deflection of the needle is read by means of a light mirror, which is attached to the needle system, a telescope and scale being employed, or the image of a slit, which is illuminated by a lamp, is thrown on a divided scale after reflection at

the mirror. The image is either produced by using a concave mirror or by placing a lens in front of the plane mirror. In both cases the angle through which the reflected beam is deflected is twice as great as the angle through which the needle is turned.

The disadvantage of the form of galvanometer described above is that if the magnetic field in the neighbourhood of the galvanometer varies, owing say to the pressure of dynamos or the like, the direction in which the needle points changes. In such situations a form of galvanometer, called the suspended-coil type, is generally used. As we shall see later, if a coil is suspended so that its plane is parallel to the lines of force of a magnetic field, then on passing a current through the coil it tends to set itself perpendicular to the lines. The suspended coil galvanometer consists of a coil which is hung so as to be able to turn between the poles of a powerful magnet, the deflection when the current to be studied is passed through the coil being read by means of a mirror attached to the coil. Since the field in which the coil hangs is a strong one the small fluctuations due to external causes produce no appreciable effect.

In galvanometers which are employed for *measuring* currents we require to be able to calculate from the dimensions of the coil the magnetic field produced by unit current at the point where the needle is suspended. Hence a coil of large radius, with the wire wound in a small groove, has to be used. Suppose that a coil of radius r containing n turns is placed with its plane in the magnetic meridian, and that a needle is suspended at the centre of the coil. When a current C is passed through the coil the field produced at the centre is $2\pi nC/r$ and is at right angles to the earth's field H. Hence the needle will take up a position parallel to the resultant of these two fields, that is, will make an angle θ with the magnetic meridian such that the couple due to the earth's field is exactly equal and opposite to that due to the coil. The couple due to the earth's field is (p. 385) $HM \sin \theta$, where M is the magnetic moment of the needle, and that due to the coil is $2\pi nMC \cos \theta/r$. Hence

The tangent galvanometer.

$$2\pi nMC \cos \theta/r = HM \sin \theta$$

or

$$C = \frac{rH}{2\pi n} \cdot \tan \theta \quad . \quad . \quad . \quad . \quad (178)$$

Thus since r, H, and n are constants, the current C is proportional to the tangent of the deflection. Hence this form of galvanometer is called a *tangent galvanometer*.

If the coil of a tangent galvanometer is mounted so that it can be rotated about a vertical axis, and the angle through which it is rotated can be read off on a horizontal divided circle, another procedure for

measuring a current can be employed. The coil is first turned till it lies in the magnetic meridian and the circle is read. The current is then passed, and the coil rotated about the vertical axis till the needle again lies in the plane of the coil. Using the same notation as before, and θ now indicating the angle through which the coil has been turned, the turning moment acting on the needle due to the earth's field and tending to bring it back into the meridian is $MH \sin \theta$ as before. The moment of the force exerted by the field of the coil, which now acts at right angles to the needle, is $2\pi nMC/r$. Hence

$$2\pi nMC/r = MH \sin \theta$$

or

$$C = \frac{Hr}{2\pi n} \sin \theta \quad . \quad . \quad . \quad . \quad (179)$$

Thus the current is proportional to the sine of the angle through which the coil is turned.

The sine galvanometer.

The usual way of performing the experiment is to turn the coil till the needle is in the plane of the coils, when the current is passing in one direction, and take the reading on the horizontal circle. The current is then reversed in direction, so that the coil has to be turned in the opposite direction. The difference between the reading of the circle when the needle is again in the plane of the coil and that obtained with the current in the original direction is twice the angle θ.

Both the tangent and sine galvanometers involve a knowledge of H at the place where the galvanometer is set up, and since the value of H will vary if masses of iron or electrical machinery is in the neighbourhood they are not suited for measuring currents under commercial conditions. A modification of the suspended-coil galvanometer is therefore generally employed, in which a coil is pivoted so as to be able to turn about an axis, and its motion about this axis is resisted by a spring similar to the hair-spring of a watch. The coil is placed between the poles of a horse-shoe magnet, and when a current is passed the coil turns till the couple due to the interaction between the coil and the field is equal to the restoring couple due to the spring. The deflections of the coil are generally read by means of a pointer which moves over a scale. The scale is calibrated by determining the readings when known currents are passed, these currents being in general determined by electrolysis (see § 166). Instruments such as the above are called *ammeters*. Since only a small current can generally be passed through the suspended coil, when such an arrangement is used to measure a large current a shunt (§ 151) is used, so that only a small *fraction* of the current actually passes through the suspended coil.

158. Force acting on a Conductor conveying a Current when placed in a Magnetic Field.—If a straight conductor, in which a current is flowing, is placed in a magnetic field, so that it is at right angles to the lines of force of the field, then, owing to the magnetic field due to the current, the distribution of the lines of force of the field will be altered. In Fig. 280 are shown the lines of force due to a conductor which is perpendicular to the plane of the paper and passes through the point A when placed in a uniform magnetic field in which

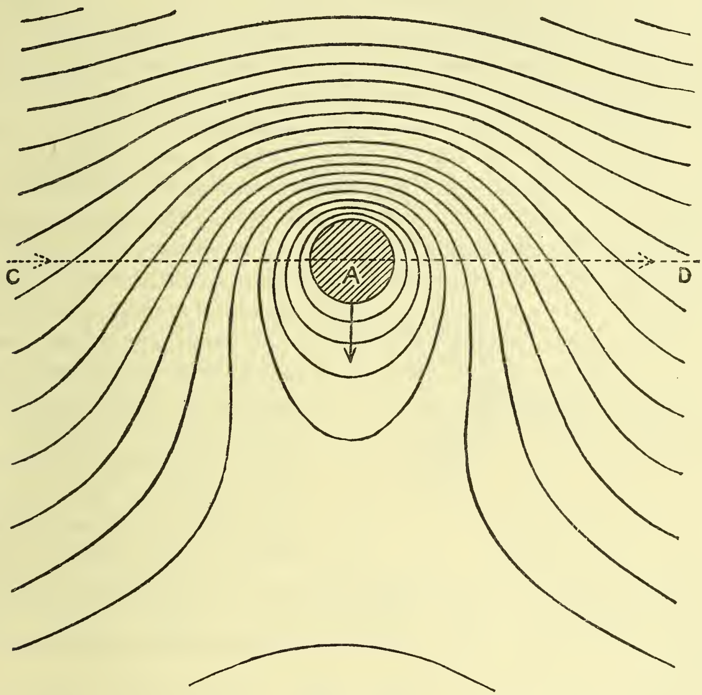


FIG. 280.

the lines of force run parallel to the line CD. Remembering that we have every reason to suppose that there exists a tension along the lines of force, and a pressure at right angles, while the lines of force act as if they were connected with the body by which they are produced, it is evident that, as a result of the crowding of the lines of force on one side of the conductor, and their separation on the other, as also to the tension along the curved lines, the conductor conveying the current will be acted upon by a force in the direction of the arrow.

If the current flows downwards, the lines of force are circles which run in the clockwise direction, and at the upper part of the diagram they

strengthen the magnetic field, since they run in the same direction as the lines of force of the field. In the lower part of the diagram the lines of force due to the current and to the field are in opposite directions, and therefore the resultant magnetic field is the difference of the fields due to the two causes. The direction in which the conductor tends to move is therefore at right angles to the direction of the lines of the field, and towards the part of the field where the lines of force due to the current are in the opposite direction to the lines of force of the field. Since the direction of the lines of force of the current can at once be remembered by one of the rules given in § 150, the direction of the force acting on a conductor in a magnetic field can at once be remembered. Fleming has given a convenient rule for remembering the direction in which a conductor conveying a current in a magnetic field will tend to move. If the index finger of the *left* hand is held pointing in the direction of the lines of force of the field, and the middle finger in the direction of the current, the conductor will tend to move in the direction of the outstretched thumb, and at right angles to the lines of force of the field. A study of Fig. 281 will make the matter clear. Thus a vertical wire in which a current was flowing downwards would, on account of the earth's horizontal component, be acted upon by a force tending to move it in an easterly direction. In this case, according to our rule, the left hand must be held with the index finger pointing towards the north, since the lines of force of the horizontal component of the earth's field run from south to north, and with the middle finger pointing downwards. The outstretched thumb will then point towards the east.

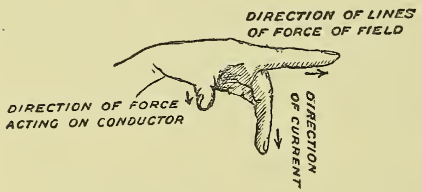


Fig. 281.

Ampère, who made a lengthy series of experiments on the forces acting on conductors in which currents are flowing, showed that if a conductor of length l is traversed by a current of C C.G.S. units, and is placed at right angles to the lines of force of a uniform magnetic field of strength H, the force acting on the conductor will be equal to lCH. If the current is measured in amperes, then, since the ampere is one-tenth of a C.G.S. unit, the force will be one-tenth of the above.

If the conductor is not at right angles to the direction of the lines of force of the field, in calculating the force we must resolve the field into two components, one perpendicular to the direction of the current, and the other parallel. Then the component parallel to the direction of the

current will produce no force on the conductor, and the force due to the other component is calculated by the formula given above.

By a consideration of the above rule it will be seen that if a current is passed through a coil which can rotate freely in a magnetic field, the coil will turn so that the maximum number of lines of force of the field thread through the coil in the same direction as the lines of force due to the current in the coil, *i.e.* so that the total number of lines passing through the coil is a maximum.

The couple tending to turn the coil will be the same as that acting on a small magnet placed at the centre of the coil with its axis perpendicular to the plane of the coil of which the moment M is given by $M=SC$, where C is the current and S is the *area* enclosed by all the turns of the coil. Thus for a circular coil of n turns, each of radius r, $S=n\pi r^2$.

Action of currents on currents.

Since a conductor in which a current flows produces a magnetic field, a second conductor conveying a current placed in this field will experience a force, and *vice versâ*. By a consideration of the direction of the lines of force due to a straight conductor, and the rule given above for the force acting on a conductor in a magnetic field, it follows that two parallel conductors will attract one another if the currents are both flowing in the same direction, and will repel one another if they are flowing in opposite directions.

The action of a fixed coil on a movable coil in both of which the same current is circulating is used, in instruments called dynamometers, to measure the current. In one form the fixed coil is vertical, and the movable coil is concentric and pivoted so that it can turn about a vertical axis. The rotation of the movable coil is resisted by a spring, so adjusted that when no current is passing the axes of the two coils are at right angles. When the current passes, the movable coil tends to set itself with its axis parallel to that of the fixed coil and takes up a position where the couple due to the action of the two coils is equal and opposite to that due to the deflection of the spring. Since, if the direction of the current in *both* coils is reversed, the couple exerted by one on the other is unaltered in magnitude or *direction*, the above arrangement can be used for measuring currents which alternate in direction. The deflection is proportional to the product of the currents flowing in the two coils, hence if a current C flows through the two coils in series the deflection varies as C^2.

CHAPTER VII

INDUCED CURRENTS

159. Faraday's Law. Lenz's Law.—In 1831 Faraday took a ring of iron and on it wound two coils of insulated wire. Having connected one of these coils to the terminals of a galvanometer, he passed an electric current through the other coil, and then found that at the moment of starting the current the needle of the galvanometer was deflected, showing that a current was passing in the second closed circuit. This deflection was only momentary, and the galvanometer immediately came back to its undeflected position, although the current in the magnetising coil was still flowing. On breaking the current, however, another momentary deflection of the galvanometer took place, but in the opposite direction to that which had occurred when the current was started.

He next wound two coils alongside one another on a wooden cylinder, and again found that when an electric current was either started or stopped in the one coil, a galvanometer connected with the other coil indicated the passage of a momentary current, the direction of the current when the main current was started being opposite to that obtained when the current was stopped.

Finally he found that if a magnet is inserted into a coil, at the instant when the magnet is inserted a current is produced in the coil, and that when the magnet is withdrawn a current in the opposite direction is also produced. He also found that if a wire, the ends of which are connected to the terminals of a galvanometer, was passed between the poles of a powerful horse-shoe magnet, so that the direction of motion of the wire was such that it cut across the lines of force of the magnet, then a current was produced during the time that the wire was being moved across the lines of force.

The currents which are produced in these ways in a closed circuit when a current in a neighbouring circuit is started or stopped, or by the relative motion of the circuit and a magnet, are called *induced currents*.

These results obtained by Faraday, which, as we shall see, are the foundation on which are based all the modern methods of producing the currents that are used in such numberless ways, such as in the production of light, the moving of vehicles, driving machinery, and performing many chemical processes, can all be summed up in the following short law:—

Whenever, from any cause whatever, the number of lines of force which thread through any conducting circuit is altered, an electromotive force will be produced during the change in the number of lines, and this will produce or tend to produce a current in the circuit.

The direction in which the induced currents flow has been put into a concise form by Lenz, in what is known as Lenz's law, and is as follows:—

The direction of the induced current produced in a conductor due to the movement of a magnet, or to that of a circuit in which a current is flowing, is always such as, by the action of the induced current on the magnet or current-conveying conductor, to produce a force tending to oppose the motion.

Lenz's law.

Thus suppose there are two parallel conductors, in one of which a current is flowing, and that the distance between the conductors is decreased, then the direction of the induced current will be such as to oppose the motion, that is, will be such as to cause repulsion between the conductors. Hence, since repulsion takes place when the currents are in opposite directions, it follows from Lenz's law that, when the conductors are moved nearer, the induced current will be in the opposite direction to the inducing current.

If, instead of the distance between the conductors being altered, the current is either started or its strength is increased, then we may look upon this as being the same thing as if the conductor in which the current is started were moved up to its present position from an infinite distance. While if the current is stopped or its strength decreased, then this is equivalent to the conductor in which it flows being removed from the neighbourhood of the conductor in which the electro-magnetic induction is produced. Hence it follows from Lenz's law that the direction of the induced current when the current is started or increased in strength is the same as when the conductors are moved nearer together, namely, in the opposite direction to that in the inducing current; while when the current is stopped the direction of the induced current is the same as that of the inducing current before it was stopped.

We will now proceed to examine different cases of the production of induced currents from Faraday's point of view as to lines of force.

In the first place, the phenomenon of the production of induced currents is said to be due to electro-magnetic induction. It is called electro-magnetic induction rather than, as is sometimes done, simply induction, for the sake of preventing confusion with the use of the term induction given in § 138.

The conductor in which the inducing current flows is called the primary conductor, or simply the primary, while the conductor in which

the induced current is produced, or at any rate in which an induced electromotive force is developed, is called the secondary.

First let us consider the case of a primary which consists of a single circle of wire, P (Fig. 282), in which a current is caused to flow by a battery, B, and in which there is a key, K, by means of which the circuit can be closed or opened, and thus the current started or stopped. Let the secondary, S, consist of a similar circle of wire, in which, if we like, we may suppose a galvanometer, G, is included. If a current in the direction of the arrow is flowing in the primary circuit, lines of magnetic force will thread through the primary in the direction shown, and some of these will also thread through the neighbouring secondary. Suppose now that the current in the primary is stopped, then all the lines due to this current will vanish. Hence the number of lines which thread through the secondary will be diminished, so that an induced current will be produced. From Lenz's law it follows that the direction of this induced current will be the same as that of the current in the primary. Now the induced current in the secondary will produce lines of force, and since the direction of the induced current is the same as that of the primary current, the direction of the lines due to the induced current will be the same as those due to the primary current. Some of these lines will thread through the primary circuit, so that the effect of the induced current is to tend to keep the number of lines which thread through the primary circuit constant, although on account of the stoppage of the primary current the number of lines tends to become less. The same effect occurs in the secondary circuit, for the lines due to the induced current, which are introduced when the primary current is broken, are such as to tend to keep the induction through the secondary constant.

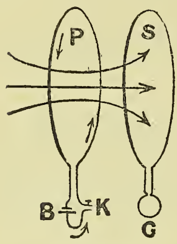


Fig. 282.

Direction of the induced current.

Next take the case where the current in the primary is started. The direction of the induced current is opposite to that of the primary current, hence the lines of force which thread through the secondary, due to the induced current, are in the opposite direction to those which are being threaded through the circuit due to the starting of the primary current. Hence the total induction through the secondary during the time the induced current lasts is the difference of the induction through this circuit due to the primary and the induced currents, so that in this case also the induced current is such that it tends to keep the total induction through the secondary circuit the same as it was before the starting of the primary current. Also, since some of the lines due to the induced current will thread through the primary, the effect of the presence of the secondary will be to postpone the time when the number

of lines through the primary reaches its final value, since their presence tends to keep the induction through the primary the same as it was before the starting of the current.

In the case when the current in the primary is kept constant, but the distance between the primary and secondary circuits is varied, the same effect takes place, namely, the induced currents are in such a direction as, by their action, to keep the number of lines of force which pass through the secondary circuit the same as it was before the motion. Of course, since the induced current only lasts while the number of lines of force is varying, the *final* induction through the secondary, as well as that through the primary itself, is quite unaltered by the fact that an induced current is produced.

Although when the secondary conductor does not form a closed circuit no induced *current* will flow in the secondary, yet in this case there will be an *electromotive force* produced owing to the electromagnetic induction, the direction of the E.M.F. being such that if the circuit were closed the current which would be produced by this E.M.F. would be that which we have been considering in the case of a closed secondary.

In the case of an unclosed secondary circuit, since there will be no induced current, there will be no lines of force due to the induced current, which, by being threaded through the primary circuit, will tend to delay the induction through this circuit from at once attaining its final value. In this case, as well as in that where there is no secondary near a circuit in which a current is started or stopped, we might expect that the current would instantly attain its final value when the circuit is closed. This, however, is not the case, for the circuit itself acts in such a way as to tend to keep the induction through itself constant. Thus before the current is started there are no lines of force passing through the circuit, but when the current is passing there are a certain number of these lines. Hence the number of lines threading through the circuit has been increased, and during the time that they were being threaded through there will be an induced current produced in the circuit itself, just as in the case of a circuit in which the increase of the total induction is due to some other circuit. As we have seen, the direction of the induced current is such as to tend to keep the number of lines of force linked through it constant. Hence when the current is started, so that the number of lines is increased, the induced lines must be in the opposite direction to those due to the current which is being started, that is, the direction of the induced current must be the opposite to that of the current which is started. The effect of this induced current, which is said to be due to *self-induction*, is to delay the current in the circuit attaining its full value, though it has no effect on the final value which

the current will reach; which final value of the current, in a simple metallic circuit, is that given by Ohm's law.

When the current in a circuit is stopped, the induced current, in order that it may tend to keep the induction through the circuit constant, must be in the same direction as the main current. The presence of the induced current when a current is stopped can be very clearly shown by means of the arrangement shown in Fig. 283. A coil, C, which ought to have a large number of turns, and an iron core (it will be remembered how the presence of an iron core increases the induction through a coil) is connected up with a make and break key, K, and a battery, B. An incandescent electric lamp, L, is connected in parallel with the coil. Although the battery may not be of sufficiently high electromotive force to cause the lamp to glow when the current is passing round the circuit, yet, when the key is opened, and induced E.M.F., due to the self-induction of the coil, will be so great that sufficient current will flow through the coil and the lamp circuit to cause the lamp to glow brightly for an instant.

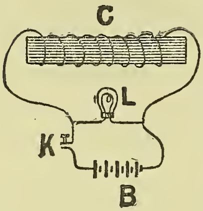


Fig. 283.

We have hitherto only considered the conditions under which induced currents are produced and the direction in which they flow, and we now have to consider on what conditions the magnitude of the induced current depends.

In the first place, the magnitude of the induced current depends on the resistance of the secondary, and since, other things remaining the same, the current is inversely proportional to the resistance, we shall in future consider the electromotive force induced in the circuit considered on account of induction; and where the value of the induced current is required, this can be calculated according to Ohm's law. The expression for the magnitude of the induced electromotive force was first given by Neumann. We may combine Neumann's results with Faraday's law as to electro-magnetic induction as follows:—

Magnitude of the induced E.M.F.

Whenever the number of lines of force which thread through a circuit is altering, an E.M.F. is produced in the circuit numerically equal to the rate at which the number of lines is diminishing.

The direction of the E.M.F. obtained by this rule is positive when it tends to produce a current in the circuit which is related to the direction of the lines of force, as the direction of rotation of a corkscrew is related to the direction of translation. Or if we are looking along the lines of force towards the circuit, then, if the number of lines is decreasing, the E.M.F. will act in the clockwise direction round the circuit.

The direction of the induced E.M.F. in a straight conductor, which is moving in a direction at right angles to the lines of force of a magnetic field, can be remembered by the following rule :—

Hold your *right* hand with the index finger pointing towards the direction in which the conductor is moving, and with the middle finger bent in the manner shown in Fig. 281, and pointing in the direction in which a north pole would travel in the field, *i.e.* so that the lines of force enter the hand at the back, then the outstretched thumb will give the direction of the induced E.M.F. in the conductor.

In a uniform field of strength F, F lines cross unit area at right angles to the direction of the lines. If then a straight conductor of length l is moved with a velocity v in a direction perpendicular both to the lines of force of the field and to the length of the wire, the space swept out by the wire in unit time will be vl. Hence the number of lines of force cut through by the wire in unit time will be $vl.F$. This then is the rate at which a circuit, of which the conductor forms a part, is increasing the number of lines of force which it embraces. Hence the electromotive force induced in the conductor due to the cutting of the lines of force of the field is vlF.

Inductance.

If when unit current circulates in a given coil L lines of force thread through the inductive area of the coil,[1] the quantity L is called the *coefficient* of *self-induction* or the *inductance* of the coil. If the current is changing at a uniform rate x, say increases by x each second, the number of lines of force threading through the coil will increase by Lx each second. Hence the induced back E.M.F. due to the self-induction of the coil will be equal to Lx. In other words, the induced E.M.F. due to the self-induction of a coil is equal to the product of the inductance into the rate at which the current in the coil is changing.

If when unit current circulates in a coil A, M lines of force due to this coil thread through the inductive area of a neighbouring coil B, M is called the coefficient of mutual induction between the coils. If unit current flows in B, then M lines of force will thread through A. If the current in either A or B is varying, the induced E.M.F. produced in the *other* will be equal to M times the rate at which the current is changing.

The practical unit of inductance (self or mutual) is called a Henry (sometimes a sec-ohm), and is such that when an ampere flows through the coil the sum of the inductances through all the turns is equal to 10^8 lines.

If the current in a coil of which the self-inductance is L henrys

[1] If N lines thread through each of the n turns of a coil when unit current passes, $L=Nn$.

is *increasing* at the rate of x amperes per second, the total induction through the coil will be increasing, and hence an E.M.F. will be produced which *opposes* the current. This back E.M.F. will be equal to Lx volts. If the current is *decreasing* at the rate of x amperes per second, an E.M.F. due to induction of Lx volts will be produced which will act in such a direction as to assist the current.

If N is the total number of lines of force which thread through a coil of which the resistance is R, and in a short time t the number of lines decreases by n, the E.M.F. produced in the coil by induction is n/t. If the ends of the coil are joined so that a current can circulate, the current produced will be n/tR. Hence the *quantity* of electricity which traverses the coil in the time t is n/R. This quantity is independent of the *time* during which the change in the induction was produced, and simply depends on this change and the resistance of the circuit. Hence the quantity of electricity which traverses a circuit owing to any change in the induction through the circuit is equal to the total change in the induction divided by the resistance of the circuit.

As an example, consider a coil of which the inductive area is A, that is, A is the sum of the areas included by all the turns of the coil, which is placed with its plane vertical and at right angles to the horizontal component H of the earth's magnetic field. The total induction through the coil is AH lines. If now the coil is rotated through 180° about a vertical axis, the total induction through the coil will still be AH lines, but it will, as far as the coil is concerned, be in the opposite direction. Hence the *change* in induction through the coil is $2AH$ lines, and if the resistance of the coil and any wires, galvanometer, or the like, included in the circuit is R, the total quantity of electricity which will circulate will be $2AH/R$.

Measurement of magnetic field by means of induced current.

The quantity of electricity which passes round a circuit due to a transient current, such as that produced when the coil is rotated, is proportional to the sine of the first swing of the galvanometer through which the current passes. The galvanometer is in this case said to be used ballistically. Hence by comparing the throws, produced when the coil is reversed (a) as above and (b) when placed horizontal, *i.e.* perpendicular to the vertical component V, we can deduce the ratio of V to H, that is, can determine the dip which is equal to the angle whose tangent is V/H.

If a copper disc is rotated in a magnetic field of which the lines of force are perpendicular to the disc, we may look upon each of the radii of the disc as a conductor which cuts through the lines of force of the field. Hence if the disc has a radius r and makes n revolutions per second, the strength of the field being H, each radius will cut through

$\pi r^2 nH$ lines per second. Thus an E.M.F. $\pi r^2 nH$ will be induced between the centre and circumference of the disc. No currents will, however, circulate in the disc. Next let us suppose that the disc A (Fig. 284) is rotated between the poles NS of a horse-shoe magnet. The portion of the disc between the centre, O, and B is cutting through the lines of force, and hence an E.M.F. will be produced tending to drive a current from O to B. In this case there is no corresponding E.M.F. induced in the remainder of the disc, and hence currents will circulate in the disc in the manner indicated by the dotted curves.

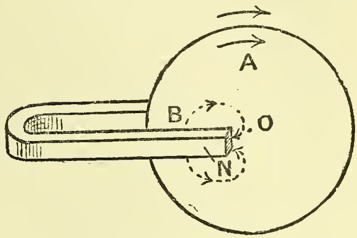


FIG. 284.

These currents which circulate in a mass of metal when it is in motion in a magnetic field are called *Foucault currents*. Since whenever a current circulates heat is produced, the disc will become heated owing to the production of the Foucault currents. The energy corresponding to this heat is derived from the work performed in rotating the disc, more work having to be done to rotate the disc between the poles of the magnet than would be required if the magnet were removed. If the disc is divided by a number of radial slits the formation of the Foucault currents is almost entirely prevented, for the whole portion of the disc between two adjacent slits is cutting the lines at very nearly

Foucault currents.

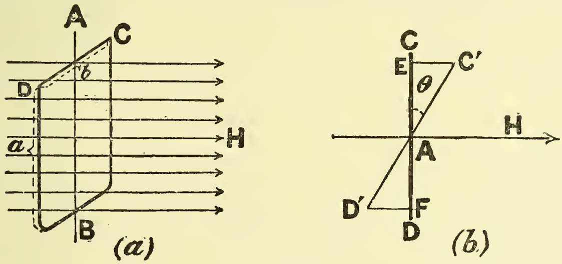


FIG. 285.

the same rate, and hence the E.M.F. induced cannot produce any current, the state of the portion of the disc in this case resembling that of the whole disc when the field was *uniform*.

160. Induced Currents produced by Rotating a Coil in a Magnetic Field.—Suppose that a rectangular circuit of length a and breadth b is rotated about an axis AB, Fig. 285 (a), which is at right angles to the lines of force of a uniform field of strength H, and that the ends of the rectangle are connected with a stationary circuit, the resistance of this circuit and of the rectangle being R. Let us start with the

rectangle in the position CD, Fig. 285 (b), in which it is at right angles to the lines of force of the field, so that the number of lines passing through the rectangle is $ab.H$. Suppose now that the rectangle is turned into the position C′D′, making an angle θ with CD. The number of lines which now pass through the rectangle is evidently equal to the apparent area of the rectangle, as seen in the direction of the lines, multiplied by H. But the area, as seen in the direction of the lines, is equal to a.EF or $2a$.EA. But EA = AC $\cos\theta = b/2.\cos\theta$. Hence the number of lines of force passing through the rectangle in its new position is $abH\cos\theta$. If the angular velocity of the coil is uniform and equal to o, and if t is the time since the coil started from the position CD, we have $\theta = ot$. Now suppose that in the small time δt the coil turns through the angle $\delta\theta$. The number of lines now passing through the circuit will be $abH\cos(\theta+\delta\theta)$. Hence in the small time δt the number of lines has decreased by $abH\{\cos(\theta+\delta\theta)-\cos\theta\}$. Now $\cos(\theta+\delta\theta)=\cos\theta\cos\delta\theta-\sin\theta\sin\delta\theta$. If $\delta\theta$ is very small, $\cos\delta\theta=1$ and $\sin\delta\theta=\delta\theta$, so that the decrease in the number of lines in the time δt is $abH\sin\theta.\delta\theta$. Now the decrease in the number of lines divided by the time during which this decrease takes place is, if the decrease goes on at a constant rate, and since δt is very small, we may consider that at any rate during this time that this is so, equal to the *rate* of decrease of the number of lines, and, as we have seen, this is equal to the induced E.M.F. Hence the induced E.M.F. is equal to $abH\delta\theta\sin\theta/\delta t$, or $abHo\sin\theta$. Hence the induced electromotive force is at any time proportional to the sine of the angle which the plane of the coil makes with a perpendicular to the lines of force of the field. Thus if E is the induced E.M.F. at the time t after the instant when the coil passed through the position CD, we have

$$E = SHo\sin\theta = SHo\sin ot,$$

where S has been written for the inductive area of the coil. There will therefore be produced a periodic E.M.F., which will vary between a maximum value of SHo in one direction, and an equal value in the opposite direction. This periodic E.M.F. will cause a periodic current to pass through a conducting circuit connected in series with the coil. Since the direction of the current will be continually reversed it is called an *alternating current.*

By suitable arrangements this alternating current in the circuit attached to the coil can be changed into a current which always flows in the same direction. Under these conditions the alternating current is said to be rectified. A method of rectifying the current consists in fitting a copper ring on the axle on which the coil turns, which is insulated from the axle, and is in addition split along two generating lines which are on opposite sides of the ring as shown at $abcd$, Fig. 286.

Two copper springs, B_1 and B_2, called brushes, rest against the copper ring, and are connected to the two ends of the external circuit. One end of the coil is connected to *ab* and the other to *cd*. The positions of the two brushes, B_1, B_2, are so arranged that as the coil revolves the brushes cross the gaps *ad* and *bc* in the ring, just as the coil is passing through the position in which its plane is perpendicular to the lines of force of the field, and hence the induced current is zero. Suppose that when the coil is in the position C′D′ (Fig. 285) the end of the coil connected with *ab* is at the higher potential, so that the current in the external circuit is going from B_1 to B_2. When the coil has passed through the position in which its plane is at right angles to the lines of force of the field, the direction of the reduced E.M.F. will be reversed; thus *dc* will now be at the higher potential. The copper conductor *dc* will now be in contact with the

Commutator.

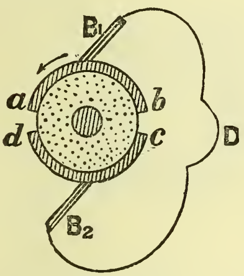


Fig. 286.

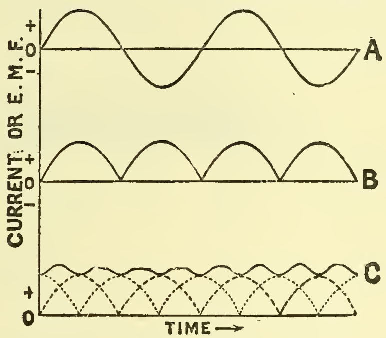


Fig. 287.

brush B_1, and hence the current in the external circuit will still flow from B_1 to B_2. Although the current in the external circuit is now always in the same direction, it is not a constant current, but twice in every revolution it is zero, and twice reaches a maximum value of oSH/R. The difference between this rectified current and the alternating current can most clearly be seen from Fig. 287, where A represents the manner in which the alternating current varies with the time, which is taken as abscissa, while at B the corresponding curve in the case of the rectified current is shown.

If a second coil of the same dimensions as the first were fixed to the same axle, so that its plane was at right angles to that of the first, and it were supplied with its own commutator, the brushes being connected to the same circuit as the first in such a way that the currents produced by the two coils in the external circuit were in the same direction, then the actual current in the circuit would be obtained by combining two such curves as that in Fig. 287 B. From the fact that one coil is placed

a quarter of a revolution in advance of the other, the two curves must be displaced by a time equal to a quarter of a revolution, the one with respect to the other. In Fig. 287 c the dotted curves represent the currents due to the two coils separately, and the full-line curve the actual current due to the combined action of the two. It will be noticed how much more nearly uniform is the current than in the case where only one coil is used, and hence it will be understood how, by increasing the number of coils, what is practically a uniform current can be obtained.

161. Dynamos.—The principles described in the last section are applied in the machines used to convert mechanical energy into electrical energy, called dynamos. These all consist essentially of a coil or series of coils which rotate in a strong magnetic field. The field is produced between the poles of an electromagnet, the whole or part of the current produced being in some cases sent round the coils of the electromagnet. This magnet is called the field magnet, and the coil or coils which rotate in the field form what is called the armature. In order to strengthen

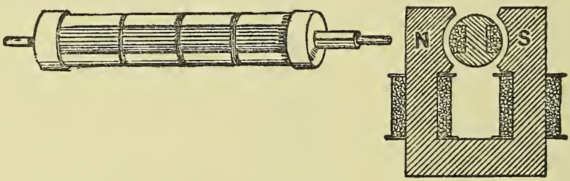


Fig. 288.

the field between the poles of the field magnet by reducing the magnetic resistance of the magnetic circuit of the field magnets, the armature coils are generally wound on an iron core, the coils either being wound on the surface of the core or in slots cut in the core. Since the direction of the magnetism of the armature core is being continually reversed it is important that the iron should exhibit as little hysteresis (§ 138) as possible. Further, since if the core is formed of a solid mass of iron, Foucault currents will be produced, it is generally built up of a large number of thin plates of iron separated from each other by an insulating varnish. These iron stampings are mounted so that the lines of force of the field do not have to cross the surfaces separating the plates, and hence the magnetic resistance of these surfaces does not affect the strength of the field. The insulating surfaces do, however, greatly reduce the circulation of the Foucault currents.

The simplest form of armature is shown in Fig. 288, and is called the Siemens shuttle armature. This armature contains only one coil, and hence even with a commutator the current fluctuates greatly, though always passing in one direction.

To avoid these fluctuations in current, armatures containing a large number of coils are employed. One type of armature, called the Gramme, containing a number of coils is shown diagrammatically in Fig. 289. This armature consists of a soft iron ring AA′ on which is wound a continuous coil of wire. The commutator used consists of a number of copper bars which are separated from one another by some insulating material, usually mica. Each of these bars is connected with a point on the wire which is wound on the iron ring. The armature is capable of being rotated about an axis perpendicular to its plane between the poles, NS, of an electromagnet. On account of the greater permeability of the iron of the ring than that of the air or other non-magnetic materials between

The Gramme armature.

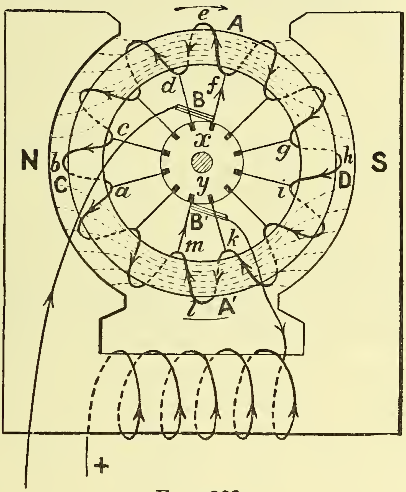


FIG. 289.

the poles, the lines of force crowd through the iron in the manner shown in the figure. Suppose now that the armature is rotated in the direction indicated by the arrow, and consider one turn of the wire *abc* which is wound on the ring. In the position in which the turn *abc* is shown there are no lines of force passing through it. As, however, the armature rotates the number of lines passing through the coil increases till it reaches the position *def*. The result of the increase of the number of lines of force passing through the coil is to cause the production of an induced E.M.F. tending to send a current in the direction shown by the arrow. As the coil passes from *def* to *ghi* the number of lines which pass through it decreases, and an induced current in the reverse direction is produced. As the coil passes from *ghi* to *klm* the number of lines which thread through it increases, but since they now pass through in

the opposite direction, the induced E.M.F. is in the same direction as it was between *def* and *ghi*. Between *klm* and *abc* the number of lines which pass through in this new direction decreases, and since this is the same thing as an increase in the number of lines passing in the opposite direction, the induced E.M.F. will be in the same direction as it was while the ring was passing from *abc* to *def*. During the rotation of the armature each coil in succession goes through the same series of conditions as the one we have been considering, and the result is that the induced E.M.F. in the half ADA′ of the coils are all in the same direction, and so the actual induced E.M.F. between the points *f* and *m* is the sum of the E.M.F.'s induced in the separate coils between these points, while an equal and opposite E.M.F. is induced in the coils in the half A′CA. Since the induced E.M.F.'s in the two halves of the armature are equal and opposite, there is no E.M.F. tending to cause a current to circulate round the armature, although this consists of a closed circuit, but an E.M.F. is produced between the bars x and y of the commutator. Hence if two brushes, B, B′, make contact with the commutator at x and y respectively, and these brushes are connected to an external circuit, a current will be produced in this circuit. If the strength of the current produced by the machine is C, then each half of the armature is traversed by a current $C/2$.

The figure also shows the manner in which the current produced is used to magnetise the field magnets. When the machine is started, on account of the residual magnetism retained by the cast iron of which the cores of the field magnets are composed, there exists a weak field between the poles. The rotation of the armature in this field produces a small current, which traverses the coils of the field magnets and increases their magnetism, and this increase in the field increases the induced E.M.F., and hence also the current passing through the field magnets. This action of the induced current in increasing the strength of the field goes on till, on account of saturation, the magnetisation of the magnets does not increase as the magnetising current increases.

Since in the Gramme armature the inside portion of each turn of the wire on the armature moves in such a way that, as shown in Fig. 289, it does not cut any lines of induction, or at any rate very few, this portion of the wire has very little beneficial effect as far as the production of an induced E.M.F. is concerned, while, since the induced current has to pass through this wire, electrical energy is wasted in heating the wire, according to Joule's law. Hence the Gramme armature is better fitted for the production of small currents at a high potential than of very strong currents. On this account a different form of winding, called the drum winding, is adopted. In one form of armature a cylindrical core of soft iron is mounted so as to be capable of rotation about its axis

between the poles of the field magnets, and on the surface of this core are arranged a number of insulated copper conductors with their length parallel to the axis, and connected together and to the commutator

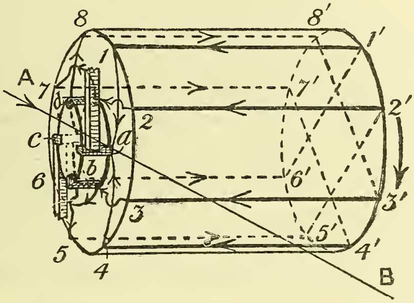


FIG. 290.

segments. There are various methods of connecting the conductors in a drum armature, one of these being indicated in Figs. 290 and 291.

The drum armature.

In Fig. 290 the conductors, 11′, 22′, &c., are shown in position on the cylindrical iron core in perspective. In order to make the system involved in the connections clearer, the armature is sometimes represented diagrammatically as in Fig. 291. In this *a*, *b*, *c*, and *d* are the sections of the commutator, while the heavy

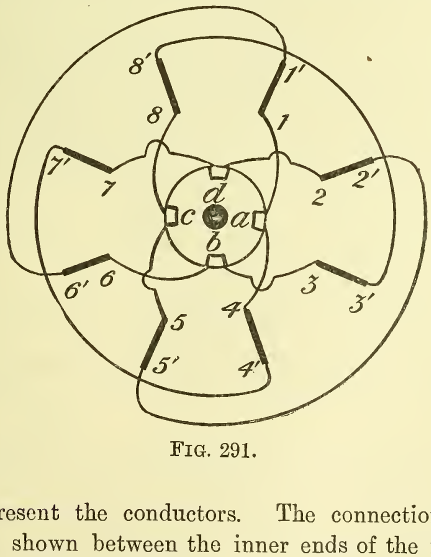


FIG. 291.

radial lines represent the conductors. The connections at the commutator end are shown between the inner ends of the radial lines, and those at the far end of the armature are shown between the outside

ends of the radial lines. It will be noticed that all the conductors are arranged in series, so that the winding forms a continuous circuit as in the case of the Gramme armature.

If in Fig. 290 the armature rotates in the direction of the arrow, while the field due to the field magnets is in the direction AB, conductors 2 and 3 will be cutting through the lines of force near the S pole, and the induced current will flow towards the commutator end of the armature, as shown by the arrow heads. In conductors 6 and 7 the induced current will flow away from the commutator. It will be seen that although conductors 1, 4, 5, and 8 are hardly cutting through any lines at the instant shown, yet they are traversed by the currents produced in the other conductors. The current produced flows away to the external circuit by a brush resting on the commutator segment *a*, and returns by a brush on segment *c*.

Series, shunt, and compound windings.

If the field-magnet coils are so connected that the whole of the current produced in the armature passes through these coils, the machine is said to be *series* wound. If the speed of the dynamo is kept constant, the voltage, *i.e.* induced E.M.F. at the brushes of the commutator, will be proportional to the magnetic field in which the armature turns. Since the strength of this field depends on the current passing in the field coils, the voltage will therefore vary as the current which the machine is called upon to supply is varied. For many purposes, particularly electric lighting, this is a disadvantage, and hence other methods of connecting the field coils are employed. In *shunt*-wound machines the field coils are connected in parallel with the armature, so that only part of the current produced is sent through the field coils. With this arrangement, if the resistance of the external circuit is reduced, so that the machine is called upon to furnish a greater current, the *fraction* of the current which goes through the field coils decreases, and as a result the voltage decreases with increase of the current. Hence a combination of the above two types of winding is often employed. In these machines, called *compound wound*, part of the magnetisation of the field coils is due to the passage of all the current in a few series windings, while the rest is due to the passage of a fraction of the current in a number of shunt turns. By suitably adjusting the numbers of the series turns and the resistance of the shunt turns the voltage may be kept constant over a considerable range of current.

If either of the dynamos described above is supplied with current from some external agency the armature will tend to rotate owing to the reaction between the field produced by the field magnets and the magnetism induced in the armature by the current which is passing. Hence they will all function as motors.

162. Back E.M.F. in Motors. Efficiency.—If a current is supplied to a direct-current dynamo it will function as a motor, and electrical energy will be converted into mechanical energy and hence do work.

Suppose that the resistance of the armature and field magnets of a motor is R, and that it is connected to a source of electromotive force, say a battery, which will produce a constant difference of potential of V volts at the terminals of the machine. Then if the armature is at rest, a current C, given by the equation $C=V/R$, will pass through the armature. If now the armature is set free, so that it is allowed to revolve, then, since if the armature were driven round in the same direction as that in which it turns an E.M.F. would be developed at its terminals in the opposite direction to that which is used to drive it, it follows that the armature by its motion will create an induced E.M.F. which will oppose the E.M.F. V which is sending a current through the motor. This counter E.M.F., as it is called, will increase as the speed of the motor increases, since the induced E.M.F. depends on the speed with which the conductors on the armature cut through the lines of force of the field. Let v be the counter E.M.F. developed at any given speed, then the effective E.M.F. sending a current through the machine is $V-v$, and hence the current which traverses the armature is given by

$$C=(V-v)/R \quad . \quad . \quad . \quad . \quad (180)$$

If the machine is supposed to turn without friction and to do no external work, the speed will go on increasing till the counter E.M.F. is equal to V. Under these conditions there will now be no force acting on the armature tending to make it rotate, and hence, since we have postulated the absence of friction, the machine will continue to turn at a constant speed. If now the machine is caused to do external work, say to wind up a weight, then the speed will decrease, and the back E.M.F. will decrease, so that a current will pass through the machine.

Suppose that the power developed by the machine, that is, the rate at which it does work, is P, and that either the friction in the different parts of the machine is so small as to be negligible, or what comes to the same thing, that the power P includes the work done against friction. If then the back E.M.F. is v, the current passing through the machine will be $(V-v)/R$. The energy corresponding to this current will be spent partly in heating the wire forming the armature, and partly in doing the work P. The part of the energy spent in producing heat is, by Joule's law, C^2R or $(V-v)^2/R$. Since the E.M.F. between the terminals of the machine is V, the energy supplied by the current C in

one second when flowing through this drop of potential is CV. Hence the energy available for doing external work is

$$\frac{(V-v)V}{R}-\frac{(V-v)^2}{R}$$

or

$$P=v(V-v)/R \quad . \quad . \quad . \quad . \quad (181)$$

From this expression it will be seen that P is zero, that is, the machine does no external work, both when $V=v$ and when $v=0$. The first case is that which we have already considered, when the motor revolves at such a speed that the back E.M.F. is equal to the applied E.M.F. The other case, when $v=0$, is when the armature is at rest, and when the current is V/R, and hence the heat developed according to Joule's law is C^2R or VC, that is, is equal to the energy supplied by the external source, so that there is none available for doing external work. If the speed of the motor is by some external means increased, so that v is greater than the applied E.M.F., the motor will operate as a generator and will send a current in the reverse direction round the circuit, and in this way will supply energy to the circuit.

The power developed by the motor will be a maximum when $V=2v$. Then the power given by the motor is $V^2/4R$, and the energy supplied will be $V^2/2R$; so that the power developed will be a maximum, when the speed is such that the back E.M.F. is half the applied E.M.F., and half the energy supplied will be converted into useful work and half wasted in heat. It must, however, be noted that although this is the speed for which, having given the external E.M.F., most work can be done by the motor, it is by no means the most economical speed at which to run the motor. The energy supplied is $V(V-v)/R$, while the energy converted into work is $v(V-v)/R$. Hence the ratio of the energy converted into useful work to the energy supplied is $v(V-v)/V(V-v)$ or v/V. Thus the proportion of the energy supplied which is converted into useful work increases as v is made more nearly equal to V. As we have seen, however, as the speed is increased, so that v may become more nearly equal to V, although the proportion of the energy supplied which is converted into useful work is large, yet, since the amount of energy which the motor is then capable of taking from the external circuit is very small, the power developed must also be small. In practice it is usual to run motors at speeds so much above that for which $V=2v$, that nearly 90 per cent. of the energy supplied is converted into useful work.

163. Alternating Currents.—As we have seen in § 160, when a coil is rotated in a magnetic field with an angular velocity o, the induced E.M.F., e, at any time t is given by $e=E \sin ot$, where E is the maxi-

mum value of the E.M.F. If the coil is connected to an external circuit and the resistance of the whole circuit is fairly high compared with the self-induction, an alternating current will be produced in the circuit, such that

$$c = C \sin ot \quad . \quad . \quad . \quad . \quad (182)$$

where c is the instantaneous value of the current at the time t and C is the maximum value.

Since, if T is the time the coil takes to complete one revolution, $o = 2\pi/T$, we have

$$c = C \sin \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad (183)$$

Although in the case of an alternating-current dynamo the relation between the E.M.F. and the time is in general not a *simple* sine curve

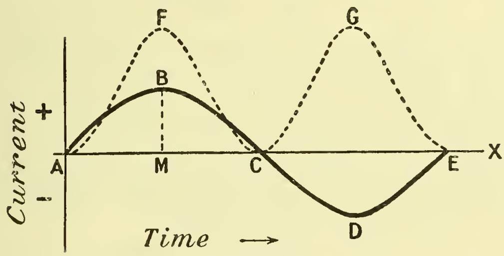


FIG. 292.

(§ 84), yet it so nearly resembles a sine curve that we may for most purposes assume that it is exactly a simple sine curve, and we now proceed to consider some of the properties of alternating currents on this assumption. If we take a sine curve ABCDE (Fig. 292), such that the length AE is equal to the period T of the alternating current, and the amplitude MB is equal to C, then the ordinates of this curve will give the values of the current c at any time throughout a complete period. An examination of the curve shows that the mean value of c is zero, for the positive loop ABC of the curve is equal to the negative loop CDE.

Since the square of a negative quantity is positive, the value of c^2 is always positive, the values of c^2 being shown by the dotted curve AFCGE.

Now since $\sin^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$, we have

$$c^2 = C^2 \sin^2 \frac{2\pi t}{T} = \frac{C^2}{2} - \frac{C^2}{2} \cos \frac{4\pi t}{T}$$

Now it can easily be shown that the mean value of $\cos \frac{4\pi t}{T}$ throughout a period is zero, for by plotting the value of this quantity we should obtain a curve similar in form to ABCDE on Fig. 292. Hence the *mean*

value of the square of the current is equal to $C^2/2$, that is half the square of the maximum value.

The square root of the mean value of the square of the current is called the *root-mean-square* of the alternating current, and is often indicated by the letters R.M.S. Thus the R.M.S. of an alternating current is equal to the maximum value of the current multiplied by $1/\sqrt{2}$ or ·707.

Root-mean-square value.

The mean value of the current for a *half* period during which the current is throughout in one direction is equal to the maximum value multiplied by $2/\pi$, or the R.M.S. value multiplied by $2\sqrt{2}/\pi$ or ·90032.

If an alternating current is passed through a resistance R the heat developed in a time t is equal to the mean of the square of the current multiplied by Rt. Hence the heat developed is the same as that which would be produced by a steady current equal to the R.M.S. of the alternating current. Thus the R.M.S. value of an alternating current is called the effective value of the current, or often simply the current. Hence when we speak of an alternating current of 100 amperes, we mean a current of which the R.M.S. is 100 amperes, and hence the *maximum* value of the current in either direction is $100\sqrt{2}$ or 141·4 amperes. Similarly the R.M.S. of the E.M.F. is called the effective voltage or potential difference, and the maximum and minimum voltages reached are $\pm\sqrt{2}$ times the effective volts.

The consideration of the methods of measuring alternating currents or voltages cannot be dealt with in the space available. We may, however, mention that the current is generally measured by the heat generated, or by an electro-dynamometer, and as both these methods measure the mean square of the current, the scales are graduated to give the R.M.S. or effective current direct. Similarly the electrostatic voltmeter described on p. 434 gives deflections proportional to the square of the volts and hence can be used to measure effective volts.

If an alternating E.M.F. of effective value $\bar{E}$ volts is applied to a circuit of resistance R, of which the self-induction is negligible, the effective current $\bar{C}$ which passes is given by $\bar{C}=\bar{E}/R$, and the maximum current passes at the instant when the P.D. is a maximum, *i.e.* the current and P.D. are in the same phase. If, however, the circuit possesses appreciable self-induction the effective current is less than the value given by the expression $\bar{C}=\bar{E}/R$, and the phase of the current lags behind that of the P.D., that is, the current reaches its maximum value after the P.D. has passed through its maximum value. If θ is the lag, the instantaneous values of the P.D. and of the current may be expressed by the equations

$$e=\sqrt{2}\bar{E}\sin\frac{2\pi t}{T} \text{ and } c=\sqrt{2}\bar{C}\sin\left(\frac{2\pi t}{T}-\theta\right) \quad . \quad (184)$$

It can be shown that the mean amount of work per second, *i.e.* the power, expended in the circuit is given by $\bar{E}\bar{C}\cos\theta$.

If there is no self-induction the power is $\bar{E}\bar{C}$, and hence $\cos\theta$ is called the *power factor* of the inductive circuit.

An alternating-current dynamo, or alternator, in which a single alternating current is produced is called a single-phase machine. The

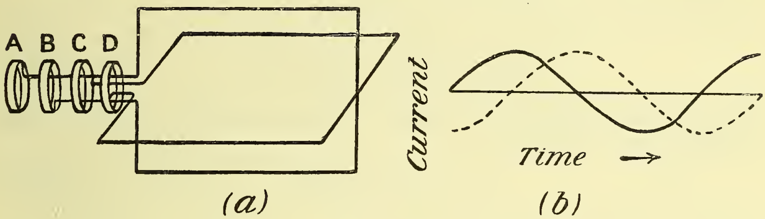


Fig. 293.

coil considered in § 160, if the ends of the wire are brought to two insulated metal rings, called slip-rings, on the axis on which two collecting brushes rest, would form a single-phase alternator. Suppose now we had two coils arranged with their planes at right angles as shown in (*a*), Fig. 293, the ends of the coils being brought to four insulated slip-

Single-phase and two-phase currents.

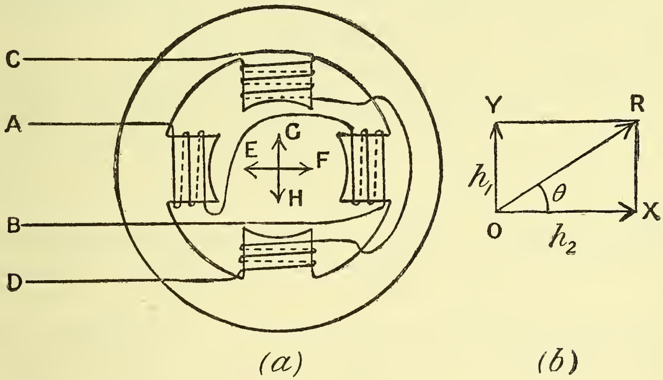


Fig. 294.

rings A, B, C, and D. Then either A and B or C and D would give an alternating current when connected to a circuit, but the phase of one alternating current would always be 90° or a quarter period behind that of the other, so that the currents in the two circuits would be related as shown by the two curves at (*b*). Such an arrangement would constitute a two-phase alternator.

Now suppose the slip-rings A and B are connected to the terminals

AB of the four-pole electromagnet shown in Fig. 294. The alternating current passing through the coils on the two opposite pole-pieces will produce an alternating magnetic field parallel to EF. Similarly, if the terminals C and D are connected to the other pair of slip-rings an alternating field will be produced parallel to GH.

Induction motor.

Let the field parallel to GH at a time t be given by

$$h_1 = H \sin \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad (185)$$

Then, since the phases of the alternating currents differ by 90°, the field parallel to EF is given by

$$h_2 = H \cos \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad (186)$$

The actual field produced will be the resultant of these two. Hence if OX, OY, Fig. 293 (b), are the values of the components at any instant, the resultant OR will make an angle θ with OX given by

$$\tan \theta = OY/OX = h_1/h_2 = \tan \frac{2\pi t}{T}$$

or

$$\theta = \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad . \quad . \quad (187)$$

Thus the value of θ changes, and OR completes one revolution in a time T, that is, in the period of the alternating currents.

The magnitude of the resultant field is given by

$$R^2 = h_1{}^2 + h_2{}^2 = H^2\left(\sin^2 \frac{2\pi t}{T} + \cos^2 \frac{2\pi t}{T^1}\right)$$

$$= H^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (188)$$

so that R is independent of the time.

Hence the resultant field is constant in magnitude and its direction rotates at a uniform rate, constituting what is called a rotating magnetic field.

If in this rotating magnetic field is placed an arrangement shown in Fig. 295, and called a squirrel-cage armature, so that the axis is perpendicular to the plane in which the field rotates, this armature will rotate, and will form an induction motor. The squirrel-cage armature consists essentially of rods of copper which are connected at either end to copper discs. If the armature were held stationary, currents would be induced in the bars by the rotating field, and by Lenz's law their direction would be such

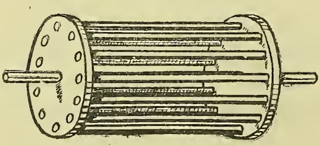


FIG 295.

as to tend to prevent the motion of the field. Hence there will exist a torque on the cage tending to rotate it in the same direction as that in which the field rotates. If the armature is free to rotate, its speed will increase till it is almost equal to that of the field, when there will be very little torque exerted. If, however, the armature is connected to a machine, say a pump, which involves the expenditure of work when rotation takes place, the armature will rotate slower than the field, so that the conductors cut through the lines of force, and thus the torque required to rotate the armature and pump is provided.

Three-phase currents.

If there are three sets of windings to an alternator, producing three alternating currents which differ in phase by 120°, the machine is called a three-phase generator, and the currents can be used in induction motors having six poles.

The currents produced by a three-phase machine are related as shown by the curves in Fig. 296. It will be noticed that at every instant the sum of the currents in the three-phases is always zero. Suppose, there-

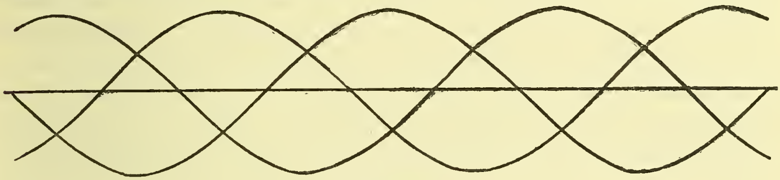


FIG. 296.

fore, we connect together one of each pair of slip-rings on the generator and one end of each of the sets of field windings on the motor, and join these by a wire X, three other wires connecting the three remaining slip-rings to the other three terminals on the motor. With this arrangement the current in the wire X will be the sum of the currents in the three-phases and hence is zero. Thus the conductor X can be done away with without interfering with the working of the arrangement, so that only three connecting wires are required between the generator and motor.

164. Transformers. Induction Coils.—The employment of electricity for the transmission of the energy developed, say, at a waterfall, to the neighbouring towns, over distances of many miles, makes the question of the cost of the conductors employed to convey the current from the generating point to the place where it is used of considerable importance. Suppose that it is required to transmit power from A to B, so that the energy available at B is W watts. If R is the resistance of the conductor extending from A to B and back, by means of which the current is conveyed, and C is the current transmitted, while V is the E.M.F. between the wires at the generating end. Then by Ohm's law the fall of potential along the wires will be equal to RC,

and hence the E.M.F. available at B will be $V-RC$. Thus the energy available at B will be $C(V-RC)$, and this is to be equal to W. The watts wasted in heat in the conducting wires is by Joule's law C^2R. Hence the object is to make C^2R as small as possible, while keeping $C(V-RC)$ constant. One way of doing this is to reduce the value of R, that is, to increase the diameter of the wire used to convey the current. This, however, involves a great outlay on copper. Another way of reducing the loss of energy in the conducting wires is to reduce the current C, but in these circumstances, if W is to remain constant, V must be made large; that is, a great potential difference between the wires must be employed. Since, however, an accidental contact with the wires conveying currents at high potentials is fatal to life, such currents are not suitable for use in houses for lighting purposes, or for driving machinery in workshops. There is a further difficulty, that to produce directly such high potential currents involves very complete insulation between the separate turns of the armature of the dynamo employed. It is thus evident that if by any means the low-tension current produced by the generator were converted into a high-tension current, and this current were transmitted to the distant station where it was again converted into a low-tension current, the advantage of the small loss of energy during the transmission with the absence of the danger attached to the use of high-tension currents would be attained. In the case of continuous currents, this transformation from low to high tension and *vice versâ* is not possible, except by virtually using a motor driven by the one current to drive a dynamo to produce the other, but with alternating currents the case is quite different.

Suppose that an iron ring (Fig. 297) is lapped over with a layer of insulated wire, there being N turns, the cross-section of the iron being s and the axial length of the ring l. Then if a current C is sent through this coil, we shall have an induction B in the iron equal to $4\pi N\mu C/l$ (§ 156). The total number of lines of force which pass through the cross-section of the iron is sB, so that the number of lines which pass round the ring of iron is given by

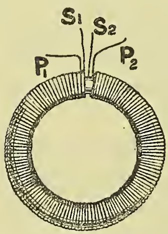


Fig. 297.

$$sB=\frac{4\pi\mu sNC}{l}$$

If in addition a second coil containing n turns of wire is lapped round the ring, then each turn of this coil will be traversed by sB lines of force, or the whole coil will be traversed by nsB lines. If now, instead of the current being constant it is an alternating current, and if C is the maximum value of the current, the induction through the

secondary coil will vary from $+nsB$ to $-nsB$, and an induced E.M.F. will be produced. If R is the resistance of the primary coil, then by making R small, that is, having a few turns of thick wire, the applied E.M.F. required to send the current C through the primary coil may be made small. On the other hand, since the induction sB takes place through each of the turns of the secondary, and that the induced E.M.F. in the whole coil is the sum of the induced E.M.F.'s in the separate turns, by making the number of turns, n, in the secondary large the induced E.M.F. may be made large. Hence by sending an alternating current through one of the coils an alternating induced current will be produced in the other coil, and by suitably varying the ratio of the number of turns in the two coils the induced E.M.F. may be made to bear any required relation to the E.M.F. used to send the current in the primary circuit. The exact relation between the primary and the secondary E.M.F.'s is complicated by the effects of self and mutual induction as well as by the hysteresis of the iron. The above will, however, explain the general principles on which transformers, as such arrangements are called, work.

In the employment of transformers for the transmission of power the generating dynamo gives a relatively low voltage, and by means of a transformer the current produced is transformed into a high-pressure current which is transmitted to the place where the electrical energy is to be used. Here, by means of a second transformer, the current is again converted into a low-pressure current.

When the secondary circuit of a well-designed transformer is not closed, owing to the self-induction, the current in the primary lags nearly 90° behind the applied P.D., hence the power factor (cos θ, p. 487) is nearly zero. Thus although an alternating P.D. of E effective volts is sending an effective current C through the primary, the power expended, which is equal to $CE \cos\theta$, is quite small. When the secondary of the transformer is closed, the induced current which is produced reacts on the primary in such a way that the phase-difference between the P.D. and current is reduced. Thus since the power factor, cos θ, increases as θ is decreased from 90°, the power absorbed by the transformer from the supply mains increases as θ decreases. This additional power absorbed is of course used up in producing the induced current in the secondary circuit.

In order to keep the power factor low when the secondary circuit is open, transformers are made with iron cores which form a closed iron magnetic circuit such as that of the ring in Fig. 297.

The secondary E.M.F. produced in a transformer depends not on the total change of the induction through the secondary, but on the *rate* of change of this induction, though, as we have seen on p. 474, the quantity

of electricity which circulates in the secondary depends on the change of induction and not on the rate of change. Hence where it is desired to obtain a very high secondary E.M.F. it is important to make the rate of change as great as possible. Thus if a steady current is passed through the primary of a transformer and then the primary circuit is suddenly broken it is possible to get a higher induced E.M.F. than if the current is reduced to zero comparatively slowly. It is found that if a condenser is joined across the point where the circuit is broken the current falls much more rapidly than if no condenser is present. The self-induction of the primary resists the change in the primary current, and causes a spark to be produced at the point where the circuit is broken, and since, as we shall see, the air in the neighbourhood of such a spark is a fairly good conductor, the current is prolonged owing to the combined action of the self-induction and the spark. The condenser, when present, is able to absorb a considerable amount of the induced current due to self-induction, so that the P.D. across the break is not sufficient to cause a spark, and hence the current is not prolonged. Since to obtain a high E.M.F. in the secondary the induction through the iron core must diminish rapidly when the primary current is interrupted, it is better to use a short straight core, for the poles developed at the end tend to demagnetise the iron (§ 138), and hence when the primary current is interrupted the core rapidly loses its magnetism.

Induction coils.

The above principles are used in the form of transformer, called a Ruhmkorff's coil, which is employed to produce sparks and to work X-ray tubes, wireless-telegraphy, and the like. The core consists of a bundle of straight iron wires on which is wound one or two layers of insulated wire to form the primary. The secondary contains a very large number of turns of fine wire, carefully insulated with silk and shellac. The current in the primary circuit being alternately made and broken, the induction through the secondary changes and an induced E.M.F. is produced in the secondary, which is in one direction when the current is made and in the opposite direction when the current is broken. Various arrangements are employed for automatically making and breaking the primary current. In some of these a small electric motor makes and breaks the current by dipping a rod of platinum into a mercury-cup. The more usual arrangement, at any rate on small coils, is to have a small piece of iron fixed to the end of a spring, so that when the current passes and magnetises the iron core the piece of iron is attracted. When no current is passing, the spring keeps the iron away from the end of the core, and makes contact between a piece of platinum fixed to the back of the iron and a platinum point which is attached to a pillar carried by the base of the coil. The primary current passes between the platinum

point and the spring, and hence when the iron hammer is attracted by the core the primary current is interrupted. The interruption of the current causes the core to lose its magnetism, so that it no longer attracts the hammer, and hence the spring forces it back against the platinum point, thus again completing the primary circuit.

165. The Electric Telegraph. The Telephone. The Microphone.—Since the direction in which a galvanometer needle is deflected depends on the direction of the current which is sent through it, by having an arrangement at one station by which the current sent by an electric battery can be reversed in direction, and connecting the commutator with a galvanometer placed at the other station by an insulated conducting wire, signals can be transmitted from the battery station to the other. Only one conducting wire is in general used, the earth being used for completing the circuit. By having a battery and a galvanometer at each station, which by means of keys can be connected to the circuit and the current reversed, messages can be sent in both directions. The older forms of electric telegraph were on this principle, the receiving instruments being, in fact, somewhat unsensitive galvanometers in which the deflection of the needle to right and left was observed. At the present time nearly all telegraphy is done by means of the Morse sounder. This instrument consists of a small electromagnet, through the coils of which the current sent from the sending station is passed. This current causes the electromagnet to attract a light soft-iron armature, which is held away from the pole of the electromagnet by means of a spring. The armature, when it strikes the pole, makes a distinct click, and from the number of clicks and the interval between them, the operator reads the signal. The current is sent by means of a key, on the depression of which the circuit of the battery is completed. In the following table the code ordinarily employed, and called the Morse code, is given. A long stroke means that the interval between that click and the next has to be longer than that between a short stroke and the next.

The Morse Alphabet.

A – ——	J – —— —— ——	S – – –
B —— – – –	K —— – ——	T ——
C —— – —— –	L – —— – –	U – – ——
D —— – –	M —— ——	V – – – ——
E –	N —— –	W – —— ——
F – – —— –	O —— —— ——	X —— – – ——
G —— —— –	P – —— —— –	Y —— – —— ——
H – – – –	Q —— —— – ——	Z —— —— – –
I – –	R – —— –	

When the distance between the sending and the receiving stations is considerable, on account of the resistance of the connecting wire, the current which can be sent is not sufficiently strong to attract the armature of the receiving instrument and make it give an audible sound. In these circumstances what is called a relay is employed. This consists of an electromagnet, round the coils of which the current which is sent from the distant station is sent. When the current passes through this electromagnet it attracts a very light and delicately poised armature. This armature works between two stops, and when it is attracted by the magnet against the one stop, it completes the circuit of a local battery which has the sounder in its circuit. Hence the sounder is worked by a battery at the station at which the signal is received, and the current transmitted from the distant station is only used to complete the circuit of the local battery.

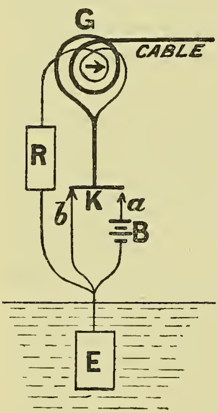


FIG. 298.

In the case of submarine telegraphy, where the distances between the stations are often very great, the receiving instrument is practically a very sensitive mirror galvanometer, and the message signals are formed by deflections of the spot of light reflected from the mirror to right and left, a deflection to the right corresponding to a dash in the Morse alphabet, and a deflection to the left to a dot.

Duplex telegraphy.

In duplex telegraphy two messages are sent simultaneously through the same wire, one in each direction. This is accomplished by winding the receiving instrument G (Fig. 298) with two coils, which are so arranged that when the battery B is connected to the circuit by pressing the key K so as to rest on the stud *a*, the current which passes from the battery divides at the instrument, part going through one coil and part through the other, and in such a direction that the effects of the currents in the two coils on the needle of the instrument are in opposite directions. One coil of the instrument is connected to the line which goes to the other station, while the second coil is connected through a variable resistance R with the other pole of the battery and the plate E, which is buried in the earth. If then the resistance of the one coil, the line, the receiving instrument at the other station, and of the return circuit through the earth, is the same as that of the second coil of the instrument at the sending station, together with the resistance R, the current which passes from B will divide into two equal parts; and since these parts traverse the coils of G in opposite directions, they will exactly neutralise each

other's effects on the instrument G, so that the working of the key K will not affect the instrument G. The current sent through the line will, however, only traverse one of the coils of the instrument at the other station, and hence it will affect this instrument.

The telephone.

The telephone was invented by Graham Bell, and a section of a modern pattern of telephone is shown in Fig. 299. It consists of a ring-shaped permanent magnet, M, to which are attached two soft-iron pole-pieces, P. Coils, A, containing a large number of turns of fine insulated wire, are placed round the pole-pieces, the ends of the wire being connected to two insulated terminals, C. A diaphragm, D, consisting of a sheet of thin wire, is clamped by its edge at a small distance above the pole-pieces, being protected by a mouthpiece, E, which is pierced by a central hole to allow the air waves produced when speaking to impinge on the diaphragm.

The diaphragm becomes magnetised by induction, and this induced

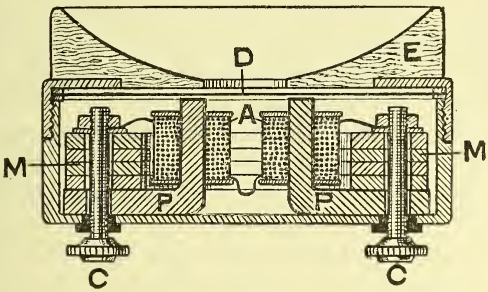


Fig. 299.

magnetisation reacts on the magnetisation of the pole-pieces P, the amount of the reaction being dependent on the distance of the centre of the diaphragm from the surface of the poles. When the instrument is spoken into, the vibrations of the air cause the diaphragm to vibrate in unison, and by its to-and-fro motion the diaphragm causes the magnetisation of the pole-pieces to vary also in unison with the incident air-vibrations. The changes of the strength of the pole-pieces mean that the number of lines of induction passing through the coil A must also vary, and hence a series of induced currents are produced in a circuit of which this coil forms a part. If the terminals C are connected by wires to a second instrument, the induced currents which are produced by the motion of the diaphragm D will traverse the coil of the second instrument, and will produce a change in the magnetisation of the pole-pieces in this instrument. These changes in the strength of the pole-pieces in the second instrument will cause changes in the force with

which they attract the diaphragm, and hence this diaphragm will be set in vibration in such a way as to reproduce the vibrations which were produced in the diaphragm of the first instrument; and in this way the air in the neighbourhood of the diaphragm will be set in vibration, and the sounds produced near the transmitting instrument will be reproduced.

The amplitude of the excursions of the telephone diaphragm is excessively small. Thus Barus has measured it and found it to be about 10^{-6} cm. when the instrument is emitting a sound which is just audible. The currents which are produced are also very small, being about 2×10^{-4} ampere in the case of the ordinary transmission of speech.

The microphone, an instrument invented by Professor Hughes, consists essentially of an arrangement by which one part of a circuit, in which is included a telephone and an electric battery, is completed by two conductors which rest lightly the one on the other. One form of the microphone is shown in Fig. 300. A piece of gas carbon, D, pointed at each end, rests lightly in two small cup-shaped hollows made in two pieces of the same kind of carbon, C, C′. The rod D is not clamped between the other rods, but rests on the lower one, and is prevented from falling by the upper end resting against the side of the cup made in the upper rod.

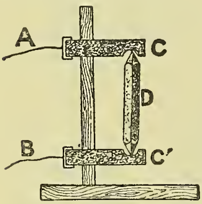


Fig. 300.

The terminals of the circuit containing the battery and telephone are attached to the rods C, C′, by the wires A and B. When a disturbance is produced, such as by the ticking of a watch placed on the base of the instrument, the rod D is set in motion, so that the pressure with which it rests against the upper rod varies. Since the resistance of carbon changes with the pressure to which it is subjected, the movements of D cause variations in the resistance of the circuit in which the microphone is included, and these changes in resistance cause corresponding changes in the current which traverses the circuit, these variations in the current causing the telephone to sound. The sensitiveness of the instrument is very great, so that even a very minute disturbance produced near the base of the instrument, such as the noise made by a fly walking on the wood, is enough to cause the telephone to reproduce the noise in quite an audible form.

The micro-phone.

The principle of the microphone has been applied to replace the telephone as a means of producing the changing currents required to transmit speech, an ordinary telephone being used to reproduce the sounds. The Blake form of microphone transmitter consists of a sheet-iron diaphragm, D (Fig. 301), held in position behind a mouthpiece by being clipped between two rubber bands. A small piece

of platinum wire, P, is attached to a slender spring, and bears at one end on the centre of the diaphragm, and at the other end against a piece of gas carbon, C, which is itself carried by a spring, B. The springs A and B are connected to a circuit in which are included a battery and the telephone at the receiving station. When the mouthpiece is spoken into the diaphragm is set in vibration, and so causes the platinum P to press on the carbon block with a variable pressure, and in this way the resistance, and hence also the current which traverses the instrument, varies in unison with the motion of the diaphragm. In this form of transmitter the energy necessary to produce the motion of the diaphragm at the distant station is supplied by the battery, and is not, as is the case when a telephone is used as transmitter, derived from the energy of motion of the receiving diaphragm. Hence in the carbon transmitter the receiving diaphragm only has to control the supply of energy of the battery, and so plays the part of the relay used in long-distance telegraphy.

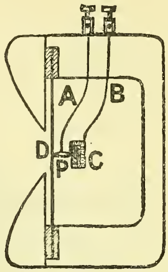


FIG. 301.

CHAPTER VIII

ELECTROLYSIS

166. Faraday's Law. Electrolytic Dissociation.—We have in the preceding pages seen that when an electric current passes through a metallic conductor a certain quantity of heat will be developed in the conductor; but after the passage of the current, except for changes caused by the rise of temperature produced by the heat developed in this way, there will be no change either in its chemical composition or physical state. In addition to metals, some liquids conduct electricity, and are called electrolytes, and we now proceed to consider what phenomena accompany the passage of a current through these bodies. The magnetic properties of a circuit which consists wholly or in part of electrolytes differ in no way from those of a circuit composed of metals only, and hence do not require any further consideration. The passage of a current through an electrolyte is accompanied, however, not only by the production of heat as in a metallic conductor, but also by certain chemical changes which take place in the electrolyte, and we now proceed to consider these in detail.

When a current is passed through an electrolyte, such as a solution of sulphuric acid in water, by dipping two platinum plates into the solution, and connecting one of these, called the *anode*, with the positive pole of a battery, and the other, called the *cathode*, with the negative pole, decomposition of the electrolyte will accompany the passage of the current. The two products of the decomposition of the electrolyte, whether they are either or both elements or compounds, will be liberated one at the cathode and the other at the anode, and not at all at any point of the liquid between. That part of the electrolyte which is liberated at the anode is called the *anion*, and that part liberated at the cathode the *cation*. It does not necessarily follow that the anion and cation are actually given off as such at the anode and cathode respectively, for secondary chemical changes often take place between the ions and the electrodes, as the plates used to form the anode and cathode are called, or with the undecomposed portion of the electrolyte. Thus in the case of the electrolysis of a solution of sulphuric acid (H_2SO_4) the cation is H, while the anion is SO_4. But while hydrogen is given off at the cathode, at the anode a secondary reaction takes place, the SO_4 reacting

with the water of the solution so as to produce sulphuric acid and free oxygen according to the equation

$$SO_4 + H_2O = H_2SO_4 + O.$$

The laws which govern electrolysis were discovered by Faraday, and are hence known as Faraday's laws. These are:—

1. The quantity of an electrolyte decomposed is proportional to the quantity of electricity which passes.

2. The mass of any ion liberated by a given quantity of electricity is proportional to the chemical equivalent weight of the ion.

In the case of elementary ions the chemical equivalent weight is the atomic weight divided by the valency, while in that of a compound ion it is the molecular weight divided by the valency. If the weight of an ion liberated by the passage of the unit quantity of electricity is called the electro-chemical equivalent of the ion, then Faraday's laws can be put into the form:—

The mass of an ion liberated is equal to the product of the quantity of electricity which passes into the electro-chemical equivalent of the ion; the electro-chemical equivalents of the ions being to one another as the chemical combining weights of these ions.

Electro-chemical equivalent.

Since, if the electro-chemical equivalent of any one ion is known, that of any other can be calculated from the chemical equivalent weights, it is of importance to determine the value of the electro-chemical equivalent in the case of one ion. Accurate experiments have shown that when one coulomb of electricity passes, that is, when a current of one ampere passes for one second, the weight of silver deposited from a solution of a silver salt is 0·001118 grams. Since the atomic weight of silver is 107·94, and the valency is 1, while the atomic weight of hydrogen is 1, and its valency is also 1, the electro-chemical equivalent of hydrogen is equal to 0·001118/107·94, or ·000010357. Hence a current of A amperes flowing for t seconds will liberate $1{\cdot}0357 \times 10^{-5} At$ grams of hydrogen, or, if q is the chemical equivalent weight of any ion, will liberate m grams of this ion where m is given by

$$m = 1{\cdot}0357 \times 10^{-5} qAt.$$

As an example, in the case of copper as a cupric salt, the atomic weight is 63, while the valency is 2; hence the chemical equivalent is 63/2, and the electro-chemical equivalent of copper is $1{\cdot}0357 \times 10^{-5} \times 31{\cdot}5$.

Since the passage of 1 coulomb will deposit ·001118 grams of silver, it will require the passage of 107·94/·001118 coulombs, or 96550 coulombs, to deposit 1 gram equivalent, that is, the chemical equivalent in grams, of silver. By Faraday's second law it follows that the passage of 96550

coulombs will cause the separation of one gram equivalent of any kind of ion. In the case of ions which can have more than one chemical valency there will be more than one chemical equivalent. Thus iron can exist in a compound either in the ferric condition, when it has a valency of 3, and consequently a chemical equivalent of 56/3 or 18·7, or as a ferrous salt, when it has a valency of 2, and hence the chemical equivalent is 56/2, or 28. Thus when a ferric salt is electrolysed the electro-chemical equivalent of iron is $1{\cdot}0357 \times 10^{-5} \times 18{\cdot}7$; while when a ferrous salt is electrolysed the electro-chemical equivalent is $1{\cdot}0357 \times 10^{-5} \times 28$.

We may suppose that the electricity passes through an electrolyte by a kind of convection, being carried by the ions, each cation carrying a definite positive charge in the direction of the current, and each anion a definite negative charge in a direction opposite to that of the current. Hence, since the passage of 96,550 coulombs liberates one gram equivalent of each of the ions, no matter what the nature of the ions, it follows that the charge carried by one gram equivalent of any kind of ion is 96,550 coulombs.[1] If we assume that the ion of hydrogen is the same thing as the atom, then one gram of hydrogen ions will have a charge of 96,550 coulombs. In the case of copper, 63/2 grams of copper ions will have a charge of 96,550 coulombs, but as this mass of copper will only contain half the number of atoms that there are in one gram equivalent of hydrogen if the copper ion is the same as the atom, each ion must have twice the charge possessed by the hydrogen ion. Thus on the above assumption as to the identity of ions and atoms the charge on each univalent ion is the same, but the charge on a divalent ion is twice as great, that on a trivalent ion three times as great, and so on. In the case of atoms which can have more than one valency there will be more than one electro-chemical equivalent, and thus ions having different charges. Thus in the ferric condition the iron ion will carry a charge three times as great as that on the hydrogen ion, and in the ferrous condition a charge only twice as great as that on the hydrogen ion.

Careful experiments have shown that in the case of electrolytes Ohm's law holds, that is, an electromotive force acting in the electrolyte produces a current which is proportional to the E.M.F. As we shall see later (§ 167), when considering the E.M.F. which is acting to produce a current through an electrolyte, it does not do to measure the E.M.F. between the electrodes by means of which the current is conveyed to and from the

[1] If we knew the number of ions in a gram of hydrogen we could calculate at once the charge carried by each ion. Although by making certain assumptions an approximate measure of the number of hydrogen ions which weigh a gram can be made, there is no particular utility in actually obtaining the number. It is, however, sometimes convenient to have a symbol to represent the charge carried by a univalent ion, and we shall use the letter *e* to represent this quantity.

electrolyte, for it requires in general a definite E.M.F. to cause a current to pass from a metal to an electrolyte, so that, when considering the connection between the E.M.F., acting between two points in the electrolyte, and the current which it produces, that is, the question of the resistance of electrolytes, the difference of potential must be measured between two points *within* the electrolyte itself.

Ohm's law being true for electrolytes, it follows that Joule's law must also be true, and hence all the energy of the current spent when traversing an electrolyte must be used simply in the production of heat, and none of it can be employed in doing chemical work in splitting up the electrolyte into ions. Within the mass of the electrolyte, therefore, the action of the current in electrolysis must simply consist in the exertion of a directive influence on the charged ions, causing them to move towards the electrodes, where, as we shall see, the work corresponding to the splitting up of the chemical compound is performed. When seeking the explanation of how, at any rate, a part of the electrolyte can be in such a condition as to allow the anions and cations to be moved in opposite directions, we are at once met with the curious fact that it has been proved experimentally that perfectly pure water is practically a non-conductor, as is also gaseous hydrochloric acid, while a solution of hydrochloric acid in water conducts freely. In the same way pure sulphuric acid is a non-conductor, or at any rate a very bad conductor, while a dilute solution of sulphuric acid is a comparatively good conductor. If we suppose that the ionisation, as the process which consists in so changing the relations of the constituents of a compound that they are able to conduct electricity is called, is due to the shaking apart of the ions in the compound molecule during the collisions between two molecules, then we should expect that the more frequent the collisions the greater the proportion of the molecules which are ionised, and hence the greater the electrical conductivity. As we have mentioned, however, this is not the case, for pure sulphuric acid does not conduct.

Electrolytic dissociation.

It is therefore evident that the hydrochloric acid and the sulphuric acid when they are dissolved in water are in a different condition from that in which they were before solution. In fact, in the solution a greater or less proportion of the molecules are either permanently split up into their ions, so that in the place of a molecule, say, of HCl we have a hydrogen ion with its positive charge $+e$, and a chlorine ion with its negative charge $-e$, or the forces which bind the H and Cl ions in the molecule of HCl are so reduced that during the collisions which occur these ions become separated, so that at any instant a finite proportion of ions are separated one from the other. It does not follow that the same ions are always thus separated, as they may recombine, but that taken as a whole there are always a

considerable number in the separate conditions. In either case the HCl is said to be dissociated or ionised by its solution in the water. In very weak solutions the electrical resistance is such that it would appear that all the molecules of the HCl are dissociated, while in stronger solutions only a fraction of the HCl molecules are dissociated, and that the undissociated molecules play no part in the conduction of the electricity in the solution.

In order to account for the dissociating influence of water, the theory has been put forward that the forces which hold the ions together to form a molecule are due to the electrical attractions between the oppositely charged ions, so that as the specific inductive capacity of water is very great, this force is very much reduced when the molecule is dissolved in water. For, as we have already seen (§ 144), if two charged bodies are transferred from air into a medium of which the specific inductive capacity is K, the force exerted between them is reduced to $1/K$th. Hence, since the ions are supposed to have a constant charge, the force exerted between the ions in a molecule, tending to prevent the splitting up of the molecule, will be less in a medium of high specific inductive capacity such as water, for which $K = 79$, than in one of small specific inductive capacity.

The dissociation or ionisation here considered is of a different nature from that which may be produced by increasing the temperature to which a pure substance is subjected, and must not be confused with it. Thus ammonium chloride (NH_4Cl) when heated dissociates into ammonia (NH_3) and hydrochloric acid (HCl), while in a solution of NH_4Cl in water dissociation takes place into the ions NH_4 and Cl.

When a difference of potential exists between two points in an electrolyte, the positively charged cations will move from the point at the higher potential towards the point at the lower, while the negatively charged anions will move in the opposite direction, the velocity with which the ions move depending on the potential gradient and the nature of the ion. Thus if the potential gradient is one volt per centimetre, chlorine ions move through a solution with a velocity of about 69×10^{-5} cm. per second. Sodium ions under the same conditions move with a velocity of 45×10^{-5} cm. per second, and hydrogen ions with a velocity of 320×10^{-5} cm. per second. These numbers are obtained by measuring the changes in concentration near the electrodes, together with the resistance of the electrolyte. The method of deducing the ionic velocities is given in the author's *Text-Book of Physics* (Longmans).

167. Polarisation.—If an electrolytic cell is prepared containing a solution of sulphuric acid, the electrodes being composed of platinum plates, and this cell and a galvanometer are included in a circuit together with a source of E.M.F., the following phenomena will occur. If the

E.M.F. of the battery is less than about 1·7 volts, on closing the circuit the galvanometer will indicate that a current passes in the circuit. The strength of the current will, however, rapidly decline, till after a short time only a very minute current will continue to pass. If the E.M.F. of the battery is greater than 1·7 volts, the current will decrease in strength for some time after the closing of the circuit, but it will never become evanescent, as was the case when the E.M.F. was below 1·7 volts. If, after the passage of a current through the electrolytic cell, the battery is removed and the circuit completed by joining together the ends of the wires which were connected to the poles of the battery, a current will pass round the circuit for some time in the opposite direction to that in which the current sent from the battery passed. Thus the plates of platinum immersed in the electrolyte possess the power not only of practically stopping the passage of a current in a circuit in which the E.M.F. is less than 1·7 volts, that is, are capable of exerting an E.M.F. in the opposite direction to that which is due to the battery, and so of preventing the passage of a current; but also this opposing E.M.F. continues for some time after the removal of the external E.M.F., so that the electrodes are able to send a current through the circuit. This phenomenon is called polarisation, and the electrodes are said to be polarised. If a current is passed through an electrolytic cell containing dilute sulphuric acid, and in which the electrodes are of platinum, and the E.M.F. between the electrodes is measured immediately after the removal of the external E.M.F., it will be found to be 1·07 volts, the anode being at the higher potential.

It must be noticed that, although according to the ionic hypothesis the ions exist in the electrolyte in the dissociated condition, it does not follow that no work has to be done to liberate the ions from the solution. In the solution each ion has its appropriate charge; when the ion is liberated at the electrode, however, this charge has been removed, so that the condition of the liberated ions is quite different from that when they were in the solution. If we assume that the hydrogen, say, as it is given off, consists of molecules each containing two atoms, these atoms being held together by chemical forces so as to form a compound, containing, however, only one kind of element; then if, as seems probable, chemical combination really consists in the holding together of the atoms by the electrical forces in play between their charges, we are led to the necessity for supposing that the molecule of hydrogen given off at the cathode must consist of a positively charged atom and a negatively charged atom. Hence in the process of electrolysis, while one hydrogen atom retains its positive charge the other loses its positive charge, and takes up from the cathode an equal negative charge, and the two combine to form a molecule of neutral hydrogen. As far as the passage of electricity through

the electrolytic cell is concerned, the giving up of its positive charge to the cathode by a hydrogen ion is exactly the same thing as taking an equal negative charge from the cathode, so that although all the hydrogen ions do not lose their charge when they are liberated, the quantity of electricity which passes through the cell, while a given number of H ions are liberated, is the same as it would be if all the H ions gave up all their charges to the cathode.

The explanation of the fact that the polarised electrodes are able, when the E.M.F. sending a current through the cell is removed, to send a current through the circuit in the reverse direction, is that the gases produced at the electrodes are not entirely liberated and given off, but that the platinum absorbs a certain quantity of the gas. On the removal of the E.M.F. this absorbed gas tends to return into the solution, and each ion of the cation, when it leaves the cathode, becomes charged positively, while each anion as it leaves the anode is negatively charged. Thus positive electricity is taken away from the cathode and negative from the anode, and hence in the external circuit connecting the electrodes a current will flow from the anode to the cathode, that is, in the reverse direction to the original current.

There are two distinct effects which are in general included under the term polarisation. One of these is the back E.M.F., which must exist when chemical decomposition is being performed by the current, in order that the requisite amount of energy may be supplied by the current. The other is an effect due to the accumulation of the products of the decomposition on or near the electrodes. Thus in the case of the electrolysis of dilute sulphuric acid between unplatinised platinum electrodes, the polarisation E.M.F. amounts to about 1·7 volts. Le Blanc has, however, shown that if platinised electrodes are employed, water may be decomposed with an E.M.F. of only 1·07 volts. When unplatinised electrodes are used the gases are evolved in the form of bubbles which form on the plates, and it would appear that a certain amount of work has to be done in producing these bubbles. With platinised electrodes, on the other hand, a much larger quantity of the gases separated by the passage of the current will not be liberated in the gaseous form, but will be absorbed by the platinum, and, in addition, the numerous small points which are present on the platinised surface seem to facilitate the evolution of the bubbles. Hence we are led to the conclusion that 1·07 volts represents the E.M.F., which corresponds to the chemical work,[1] that is, the splitting up of the chemical compound that forms the electrolyte, which is done in the cell, while the greater value, 1·7 volts, which is necessary to produce decomposition when unplatinised platinum

[1] We shall return to the subject of the connection between the E.M.F. and the energy required to perform the chemical work in § 170.

electrodes are employed, is due to the additional work which has to be done at the electrodes in performing the mechanical work of separating the liberated gases from the electrodes.

The term polarisation is also used in yet another sense, when the increase of the resistance of an electrolytic cell, due to the fact that the bubbles of gas which adhere to the electrodes virtually diminish the cross-section of the liquid conductor, and hence increase the resistance, is spoken of as being due to polarisation.

It is better, however, to call such an effect as this a transition or secondary resistance. The magnitude of these mechanical effects depends on the current density, that is, the quotient of the current passing by the area of the electrode, on the solubility of the gases liberated in the electrolyte, and on the occlusion of the gases by the electrodes. So that if the term polarisation is taken to include all these effects its value is most indefinite.

In the case of the electrolysis of a solution of copper sulphate between electrodes of copper we have copper deposited on the cathode, and the SO_4 ion attacks the anode producing copper sulphate. Hence in such a case the energy which is required to split up the salt is regained by the formation of an equal mass of the salt at the anode, so that no work has to be done by the current in producing chemical energy of separation. We should therefore expect that in this case there would be no counter E.M.F. of polarisation, which as a matter of fact is found to be the case. Similarly, when a solution of zinc sulphate is electrolysed between electrodes of zinc there is no polarisation. When there is no polarisation, the difference of potential between the electrodes during the passage of a current is equal to the product of the current into the resistance between the electrodes according to Ohm's law. If, however, there is an opposing E.M.F. of polarisation e developed when a current C is passed, the resistance of the electrolyte between the electrodes being R, then by Ohm's law there will be a difference of potential between the portions of the electrolyte near the anode and cathode respectively given by RC. In this case the difference of potential between the *electrodes* will be given by $E = RC + e$, so that $C = (E - e)/R$. Hence, when we are considering the passage of a current between the electrodes in a cell where polarisation occurs, the applied E.M.F. must be reduced by the E.M.F. of polarisation in order to calculate the current, according to Ohm's law, from the resistance of the cell.

When measuring the resistance of an electrolyte the effects of polarisation must be guarded against, for the drop of potential between the electrodes when a current C is passing is $RC + e$, and hence unless e is zero it is not sufficient simply to compare this drop of potential with that between the ends of a known resistance traversed by the same

current, as is done in the Wheatstone's bridge. Since the magnitude of the E.M.F., e, due to polarisation depends on the quantity of electricity which has passed, if an alternating current is employed the value of e can be kept very small, for the current passing alternately first in one direction and then in the other, no accumulation of the products of decomposition on the electrodes takes place. Thus one way of measuring the resistance of an electrolyte is to reduce the polarisation to a minimum by using suitable electrodes and then measure the resistance with a Wheatstone's bridge, using an alternating current. In order to supply this alternating current a small induction coil is employed, the secondary of the coil being joined to the terminals of the bridge to which the battery is usually connected. Since, when an alternating current is employed, the current which traverses the galvanometer branch when the bridge is not balanced is also alternating, an ordinary galvanometer cannot be employed; the reason being that although the bridge may not be balanced, yet the galvanometer would be undeflected, for it would be traversed by a current first in one direction and then in the other. Hence, since the period of the galvanometer needle has to be much greater than the period of the alternations, before the needle has time to move appreciably, under the influence of the current in one direction, the direction of the current is reversed, and thus the motion is checked. The usual arrangement is to replace the galvanometer by a telephone, and to move the point of contact on the slide wire of the bridge till the telephone is silent. When this adjustment is made, no alternating current is passing through the telephone, and the ordinary relation between the resistances of the bridge holds. In order that complete silence may be secured, it is important that none of the resistances forming the bridge should have either self or mutual induction. This method of measuring the resistance of electrolytes is known as Kohlrausch's method.

Measurement of the resistance of an electrolyte.

168. Contact Difference of Potential. The Voltaic Cell.—If a piece of zinc is placed in contact with a piece of copper, owing to the contact of the dissimilar metals, a difference of potential will be developed, the zinc being at the higher potential. The difference of potential is very small, being only about ·0007 volt. When, however, a metal is in contact with an electrolyte, contact differences of potential of a considerable magnitude may be produced. Thus with zinc in contact with dilute sulphuric acid there is a contact difference of potential of ·62 volt, the zinc being at the lower potential. Copper in dilute sulphuric acid is at a potential of ·46 volt higher than the solution.

Suppose now we have a plate of copper and a plate of zinc dipping

in a vessel containing a dilute solution of sulphuric acid. If we take the potential of the solution as zero, the potential of the copper will be + ·46 volt, and that of the zinc − ·62 volt. Hence the copper will be at a potential of 1·08 volts higher than that of the zinc, so that if the copper and zinc plates are joined by a conducting wire, this wire will be traversed by an electric current. This arrangement is called a simple galvanic or voltaic cell or battery.

Simple voltaic cell.

When the copper plate is connected to the zinc by the conducting wire, positive electricity will flow from the copper to the zinc, and so the potential of the copper and therefore also of the solution in its immediate vicinity will decrease. Similarly the potential of the zinc and the solution in its vicinity will increase. Thus a potential gradient will be produced in the solution, and hence electrolysis will take place, the positively charged H ions travelling towards the copper and the negatively charged SH_4 ions travelling towards the zinc. The SO_4 ions attack the

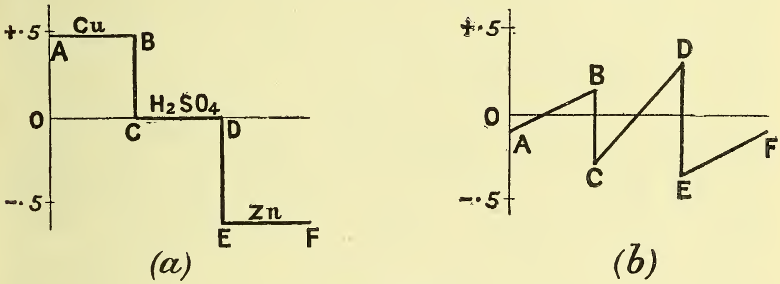


Fig. 302.

zinc, forming $ZnSO_4$, and the hydrogen is liberated at the copper plate. Thus the passage of a current in the wire joining the plates is accompanied by electrolysis of the solution, and for every equivalent of the zinc dissolved an equivalent of hydrogen will be liberated, and the passage of 96,550 coulombs of electricity in the wire will be accompanied by the solution of the equivalent of zinc, and the liberation of the equivalent of hydrogen. Thus the cell behaves, as far as the chemical changes which take place within it are concerned, exactly as if it were an electrolytic cell through which a current is sent by some external agency.

The potentials of the different portions of the simple cell, considered before the external circuit was closed, may be represented by the curve (a), Fig. 302, in which the ordinates represent the potentials of the different portions of the circuit. It will be seen that the copper, the zinc, and the solution are each at the same potential throughout. When the external circuit is closed, so that a current flows in the external wire, the

end A of the copper wire being in contact with the end F of the zinc, these two points will be at the same potential. We are here neglecting the contact difference of potential between the copper and the zinc, since this is so very small compared with the contact differences between the metals and the solution. Since there is a current flowing in the copper wire in the direction BA, there will, by Ohm's law, be a fall of potential along the wire as indicated by the line BA, Fig. 302 (*b*). In the same way, as a current is passing through the zinc in the direction FE, there will be a fall of potential along this wire as shown by FE. Similarly, since Ohm's law applies to the electrolyte, there will be a fall of potential in the liquid between D and C. Study of the figure will show that, although the copper plate is still at a potential of 0·46 volt above that of the solution in its immediate neighbourhood, as represented by BC, and that of the zinc plate is at 0·62 volt below that of the electrolyte in its immediate neighbourhood, as represented by DE, yet, owing to the fall of potential along the electrolyte so that the end D is at a higher potential than the end C, the difference of potential between the copper and zinc plates, that is, between the poles of the cell, is less when a current is passing than when the cell is on open circuit. The change produced on this account can immediately be calculated, for if r is the resistance of the electrolyte between the copper and zinc plates, that is, the resistance of the cell, then, when a current C is passing, the fall of potential will by Ohm's law be equal to rC. Hence the potential between the poles of the cell, when it is producing a current C, will be $1{\cdot}08 - rC$.

The above discussion, in the case of such a cell as the one described, only applies to the first moment of closing the circuit, for after an appreciable current has passed, polarisation effects will occur which will decrease the available E.M.F. The polarisation with which we are here concerned is of the second kind considered in § 167, that is, it is due to the effect of the hydrogen liberated at the copper plate and to a less extent due to the dissolved zinc ions not diffusing away from the zinc, so that the rate of solution of the zinc will be diminished owing to the presence of an increased number of zinc ions in the solution near the plate. The polarisation due to the hydrogen is produced, in the first place, by the copper becoming coated with a layer of fine gas-bubbles which increases the resistance of the cell, owing to the diminished surface of the copper available for the passage of the current; and, in the second place, by the copper becoming coated with hydrogen, the positive plate becomes practically a plate of hydrogen, and, since hydrogen has a smaller contact difference of potential with the solution than copper, the E.M.F. of the cell is decreased. A cell such as the above, in which, owing to polarisation, the E.M.F. decreases rapidly when a current is allowed to pass, is called an inconstant cell. In order to obviate the

polarisation, we must so choose our electrolyte that the chemical processes which take place when a current passes do not cause the accumulation of the ions on the electrodes in such a way as to increase the resistance or decrease the contact differences of potential between the electrodes and the electrolyte. Cells which fulfil this condition more or less completely, and are, at any rate at first, free from polarisation, and of constant E.M.F. even after sending a current, are called constant cells, and we shall now proceed to describe some of the commoner forms of such cells. It must be remembered that there must always be a decrease in the difference of potential between the poles numerically equal to rC, where r is the resistance of the cell and C the current which is passing. This apparent reduction in the E.M.F. is due to the fact that the liquid of the cell has appreciable resistance, and hence, by Ohm's law, a certain proportion of the available E.M.F. has to be employed in driving the current through the liquid, that is, in moving the H ions, in the simple cell previously considered, to the copper pole, and the SO_4 ions to the zinc pole. This effect being quite independent of the polarisation, will not be affected by any change calculated to diminish the polarisation, and can only be reduced to a minimum by making the resistance of the liquid, that is, r, a small as possible.

169. Primary Cells. The Storage Cell.—The Daniell cell consists of a zinc plate dipping in a solution of sulphuric acid or zinc sulphate and a copper plate in a solution of copper sulphate, the two solutions being prevented from mixing by the interposition of a partition composed of porous earthenware. In some forms of the cell called gravity Daniells, the porous partition is done away with, the copper plate Cu (Fig. 303) being placed at the bottom of a glass vessel and covered with a saturated solution of copper sulphate. The zinc sulphate, which has a less density than the copper sulphate solution, floats on the top of the latter, and, since convection currents cannot be formed and the process of diffusion is very slow, the solutions do not mix for some time. The negative plate is formed by a horizontal disc of zinc, Zn. The connection with the copper plate is made by means of a wire, A, which passes down to the copper and is enclosed in an insulating tube, generally of glass.

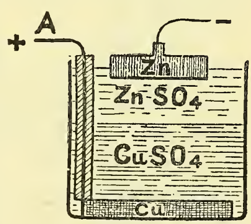


FIG. 303.

The E.M.F. of a Daniell cell is about 1·096 volts, increasing slightly with increase in the concentration of the copper sulphate solution, and with decrease in the strength of the zinc sulphate solution.

When the external circuit of a Daniell cell is closed, so that a current passes, the zinc goes into solution as zinc sulphate, while the cation of

the copper sulphate solution, that is, the copper, is deposited as metallic copper on the copper plate of the cell. It will thus be seen, since the deposition of copper on the copper cathode will in no way affect either the resistance of the cell or the contact difference of potential between the copper sulphate solution and the copper, that the E.M.F. of this form of cell will not be decreased on account of polarisation. When the cell sends a current the SO_4 ions, each carrying a charge of $-2e$, move from the neighbourhood of the copper pole to the zinc pole, and the copper ions which are left at the copper pole are deposited, giving up their positive charge. At the same time the zinc ions enter the solution from the zinc pole, each carrying a positive charge, and these positive ions, together with the negative SO_4 ions which have migrated from near the copper pole, being in equivalent proportions in the solution, prevent the solution becoming charged.

In the Grove cell the positive pole consists of a plate of platinum in a strong solution of nitric acid, and the negative pole is a zinc plate in a fairly strong solution of sulphuric acid (1 of acid to 10 of water), the liquids being separated by a porous earthenware partition. The E.M.F. of this cell is about 1·97 volts. When a current passes, the zinc goes into solution, forming zinc sulphate with the SO_4 ions of the sulphuric acid solution; while the H ions migrate, each carrying its positive charge, to the platinum plate, where they give up their charge and thus transport the current through the cell. The hydrogen is not, however, given off at the platinum, but a secondary reaction takes place between it and the nitric acid, which results in the combination of the hydrogen with part of the oxygen of the acid to form water, and leaves an oxide of nitrogen in the solution. The E.M.F. of the cell gradually falls off, owing to the exhaustion of the nitric acid allowing polarisation to take place, as well as to the gradually increasing concentration of the zinc ions in the solution diminishing the potential fall from the acid to the zinc.

The Bunsen cell is the same as the Grove cell, except that the positive pole consists of a plate of gas carbon. A solution of chromic acid is sometimes used in place of the nitric acid, the action being of a similar nature. Since the presence of chromic acid near the zinc does not materially affect the solution of the zinc ions, and does not produce any secondary chemical action with the zinc, the porous cell separating the chromic acid and the sulphuric acid solution may be omitted. This form of cell is called the chromic acid cell, or, since bichromate of potash is sometimes used in place of the chromic acid, the bichromate cell.

The positive pole of the Leclanché cell consists of a plate of carbon packed round with a mixture of powdered carbon and manganese dioxide

(MnO_2). The negative pole consists of zinc, and the electrolyte is a solution of ammonium chloride or sal-ammoniac (NH_4Cl). During the passage of the current Cl ions move towards the zinc, forming zinc chloride, while the NH_4 ions move towards the carbon cathode, where they break up into ammonia (NH_3), which dissolves in the solution, and hydrogen, which combines with part of the oxygen of the MnO_2 to form water and MnO. The E.M.F. of this cell on open circuit is about 1·6 volts, and with any but small currents the cell polarises rather rapidly. After a short rest, however, the cell recovers, and it possesses the great advantage that no chemical action goes on when no current is passing, and as there is only one kind of electrolyte, diffusion of one electrolyte into the other, which always occurs in time when two electrolytes are employed, does not occur.

A form of cell often used when a strong current for a comparatively short time is required, consists of a carbon plate in a mixture of dilute sulphuric acid and potassium or sodium bichromate, a zinc plate in sulphuric acid forming the negative pole, the two solutions being separated by a porous pot. The bichromate acts as a depolariser, as it gives up oxygen to the hydrogen ions so that they are not deposited in the carbon.

The above described forms of primary cells are intended to provide a current, and the fact that their E.M.F. varies slightly from time to time is of little consequence. We now proceed to describe a cell which is not intended to provide any but a very minute current, but of which the E.M.F. is very constant, so that the cell can be used as a standard of E.M.F. The cell is called the cadmium cell, and consists of an H-shaped glass tube (Fig. 303), platinum wires being fused through the bottom of the limbs. The positive pole, B, consists of some pure mercury, while the negative pole, A, consists of an amalgam of cadmium. Over the mercury is a layer of paste formed of mercurous sulphate and a saturated solution of cadmium sulphate. The rest of the tube is filled with a saturated solution of cadmium sulphate together with crystals of this salt, so that even if the temperature rises the solution will remain saturated. The E.M.F. of the cadmium cell at 20° C. is 1·0183 volts, decreasing slightly with rise in temperature, and remains constant for years.

Standard of E.M.F.

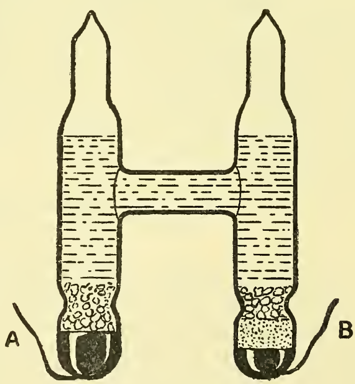


Fig. 304.

If a current is passed through a Daniell cell by means of an external

source, so as to enter the cell at the copper pole and leave by the zinc pole, the chemical reactions which will take place will be exactly the reverse of those that occur when the cell itself sends a current, for the copper will be dissolved to form copper sulphate, and the zinc deposited from the zinc sulphate solution. Hence, passing an electric current in the reverse direction through a Daniell cell of the type Zn, $ZnSO_4$, $CuSO_4$, Cu, is accompanied by chemical changes such that, when the cell is itself allowed to send a current, the inverse chemical changes take place. Thus part at any rate of the energy spent in sending the original current through the cell is stored up in such a way that it may at a future time be reconverted into electrical energy. Any other form of cell may be used in the same way, provided the products of the chemical actions which take place during the working of the cell are retained either on the electrodes or in the electrolyte, and are not given off. The simple voltaic cell, consisting of a plate of copper and one of zinc in dilute sulphuric acid, is not reversible, since the hydrogen which is evolved at the cathode when the cell is sending a current escapes in the gaseous form. A form of cell which is specially designed to store up electrical energy, so that it can be recovered at a subsequent time in the form of a current, is called a storage cell, secondary cell, or accumulator.

Reversibility of primary cells.

The commonest form of storage cell consists of two lead grids, the interstices being filled with lead sulphate formed by making a paste with one of the oxides of lead, litharge or red lead, and dilute sulphuric acid. These plates are immersed in a dilute solution of sulphuric acid, and then a current is passed through the cell from one plate to the other. During the passage of the current, the hydrogen ions of the sulphuric acid travel to the cathode, where they react on the lead sulphate, forming sulphuric acid and metallic lead, which remains in the interstices of the plate in a very spongy condition. The SO_4 ions travel to the anode, where they also react on the lead sulphate, forming peroxide of lead and sulphuric acid according to the equation

The storage cell.

$$PbSO_4 + SO_4 + 2H_2O = PbO_2 + 2H_2SO_4.$$

The peroxide of lead is left in the interstices of the grid.

When nearly, if not quite all, the lead sulphate on the grids has been changed in this way, the hydrogen ions will be liberated at the cathode in the form of gas, while at the anode, owing to the secondary reaction between the SO_4 ions and the water of the solution, which has already been referred to when considering the electrolysis of dilute sulphuric acid between platinum electrodes, oxygen is liberated. When this evolution

of gases occurs, the cell is no longer working in a reversible manner, and it has received the maximum charge of which it is capable.

If, after being charged in this way, the plates are connected by a conducting wire, a current will be obtained in the reverse direction to that employed to charge the cell, the chemical changes taking place in the reverse direction, the spongy metallic lead becoming converted into the sulphate, and the peroxide also forming the same compound.

The E.M.F. of a freshly charged accumulator is about 2·1 volts, which gradually falls to about 1·8 volts as the discharge goes on. The lead accumulator is a wonderfully efficient means of storing energy, since about 80 per cent. of the energy spent in charging the cell is recoverable if the discharge takes place within a fairly short interval after the charge. The disadvantage of the lead storage cell lies in the fact that, owing to the considerable changes in volume which take place in the active material during charge and discharge, the grids disintegrate pretty rapidly, and hence the expense of the renewals of these plates has to be taken into account when considering the efficiency of the cells from a practical standpoint.

170. Source of the Energy of the Current given by a Voltaic Cell.—We have considered the question as to how the E.M.F. of a voltaic cell is produced, and now we have to consider more in detail from whence the energy necessary for the *maintenance* of a current is derived. This energy is evidently, in part at any rate, derived from the energy of the chemical processes which take place in the cell during the time when it is sending a current. In § 75 we have considered the energy which is liberated or absorbed during certain chemical changes, and the question arises as to the connection between the total quantity of energy which is evolved as heat, when the reaction takes place without the production of an electric current, and the energy represented by the current when this is produced. It was thought for some time that the whole of the energy corresponding to any chemical change was converted into electrical energy when the change took place in a voltaic cell, and the fact that the E.M.F. of the Daniell cell, when calculated on this hypothesis from the thermo-chemical data for the chemical changes which take place in this cell, agreed very well with the value as obtained by direct measurement, supported this view.

Thus in § 75 we have seen that when 65 grams (one gram atom) of zinc are dissolved in dilute sulphuric acid according to the equation $Zn + H_2SO_4 = ZnSO_4 + H_2$, 38,066 calories are evolved. Experiment has shown that when 63 grams of copper are converted into copper sulphate in solution in water according to the equation $Cu + H_2SO_4 = CuSO_4 + H_2$, 12,500 calories are absorbed, so that 12,500 calories are evolved when $CuSO_4$ is split up into Cu and H_2SO_4. Hence when one equivalent,

that is, since zinc is a diad, 65/2 grams of zinc are converted into the sulphate, while at the same time one equivalent of copper (63/2 grams) is deposited from the sulphate, $19033 + 6250 = 25283$ calories are on the whole evolved.

Now the reactions considered above are those which go on in the Daniell cell when it is sending a current, and we have seen that the quantities of zinc and copper considered above are dissolved and precipitated respectively when 96,550 coulombs of electricity pass through the cell. If the E.M.F. of the cell is E volts, then the passage of 96,550 coulombs of electricity will correspond to $96550 \times 10^7 \times E$ ergs, for one volt is equal to 10^8 C.G.S. units, and one coulomb is 10^{-1} C.G.S. units. If, then, the whole of the energy corresponding to the chemical reaction is converted into electrical energy, we shall have, since 25,283 calories is equal to $25283 \times 4 \cdot 2 \times 10^7$ ergs, or 1068×10^9 ergs,

$$96550 \times 10^7 E = 1068 \times 10^9 \text{ ergs}$$

$$E = \frac{106800}{96550}$$

$$= 1 \cdot 106 \text{ volts.}$$

Now direct measurement has given the value 1·096 volts for the E.M.F. of a Daniell cell, so that in this case it would seem that the electrical energy of the cell is equal to the chemical energy corresponding to the reactions which go on in the cell during the passage of the current.

When, however, the same method of calculation came to be applied to other forms of cells it was found that the E.M.F.'s calculated on this hypothesis differed from the observed values by more than could be accounted for by errors of experiment. The reason for these differences was shown by Helmholtz to be due to the fact that the hypothesis that the electrical and chemical energies were in all cases exactly equal was not true. He showed that this was only true in the case of cells in which the E.M.F. does not vary with the temperature, the Daniell being a cell of this kind.

If the E.M.F. of a cell increases with increase of temperature, then when the cell is sending a current its temperature will tend to fall, and hence if its temperature remains constant heat must be *supplied* to it by conduction from surrounding objects. The electrical energy in this case will represent not only the thermal energy corresponding to the chemical changes which take place, but also this additional heat. If the E.M.F. of the cell decreases with increase of temperature, *heat* is given out to surrounding objects and the electrical energy produced is less than the thermal energy corresponding to the chemical changes.

CHAPTER IX

PASSAGE OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY

171. Passage of Electricity through a Gas. Cathode, Röntgen, and Canal Rays.—In its ordinary condition a gas is an extremely bad conductor of electricity, in fact for small P.D.'s the conductivity is practically zero. If, however, the P.D. acting between two knobs, say in air, is increased to a high value, the dielectric stress produced in the space between the knobs becomes so great that the air breaks down and a spark discharge passes between the knobs. After the passage of a spark the air in the track of the spark is found to be a fairly good conductor of electricity, so that if a comparatively small P.D. is maintained between the knobs after the passage of the first

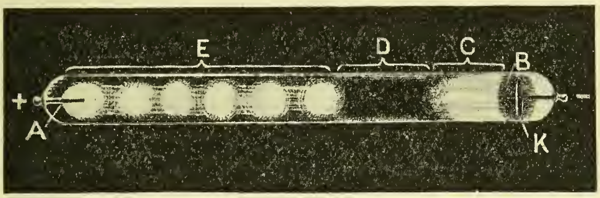


FIG. 305.

spark a current will continue to flow through the air. The air when in the conducting condition is said to be *ionised*. This ionisation persists for some time, and if a current of air is blown between the knobs and a succession of sparks is passed, say by connecting the knobs to the terminals of an induction coil, the ionised air produced is carried to one side and the succeeding spark, in place of passing straight across the shortest path between the knobs, follows the ionised air. Thus while the first spark is straight, the second spark is curved, the concave side being turned towards the current of air.

If the pressure of a gas is reduced, the P.D. required to produce a discharge and the character of the discharge changes in a very marked manner. Thus if we take a glass tube such as is shown in Fig. 305,

with platinum wires fused through the ends, these wires being connected with a small aluminium plate K and a wire A respectively, and connect the wires to a source capable of producing a high P.D., the wire A being positive, and gradually exhaust the tube, the following series of phenomena will be observed. At first, unless the P.D. is extremely high, no discharge will pass, but when the pressure is reduced to about 8 cm. of mercury a line of violet light will be seen to stretch down the axis of the tube, showing that a discharge is passing. When the pressure is reduced to about half a millimetre the appearance of the discharge will have entirely changed. At the plate K, called the cathode, there will be seen a soft glow which moves about over the surface of the cathode. Next to the cathode there is a space, B, which is comparatively free from luminosity, and which is called Crookes's space, or the first dark space. The distance from the cathode through which this dark space stretches increases as the exhaustion of the gas increases. The termination of the dark space nearest the anode is quite sharp, and is very approximately the surface on which would lie the ends of equal normals drawn from the surface of the cathode. Beyond the dark space is a luminous space C, called the negative column. The position of the negative column does not depend on that of the anode, so that if the anode is placed in a side tube the negative column does not bend round into the side tube, but goes straight on and fills the portion of the tube beyond the point where the side tube containing the anode leaves the main tube.

Beyond the negative column there is a second comparatively dark space D, called the second negative dark space. This dark space varies very much in size, and may sometimes be entirely absent. Beyond this dark space there is another luminous column, E, which extends up to the anode, and is called the positive column. The luminosity of the positive column is often not continuous, but consists of alternate bands of bright light and comparatively dark spaces. These bright bands are called striæ, and often present a very striking appearance. The colour of the striæ depends on the nature of the gas within the tube, and, according to Crookes, when a mixture of gases exists within the tube each gas produces a separate series of striæ.

When the exhaustion within a tube is carried to below a thousandth of a millimetre of mercury, the positive column gradually vanishes, and the sides of the tube exhibit a brilliant phosphorescent glow. The colour of this glow depends on the nature of the glass; thus with lead, or English glass, the glow is blue, while with German, or soda glass, the phosphorescence is of a beautiful emerald green. The appearance presented is as if something were projected by the cathode in a direction normal to its surface which, when it strikes the glass, has the power of exciting phosphorescence.

Whatever the nature of the emanation from the cathode, or the *cathode rays* as they are called, they under ordinary conditions proceed in straight lines, so that if a screen is placed between the cathode and the sides of the tube a clear shadow of the screen will be produced on the walls.

Cathode rays.

If a concave cathode is used the rays are brought to a focus at the centre of curvature of the cathode, and if a body is placed at this focus it will become heated. If the discharge is sufficiently strong even platinum can be melted in this way. Other bodies, besides glass, phosphoresce when exposed to the cathode rays. Thus rubies give out a brilliant red light.

The cathode rays are deflected when a magnet is brought near in the same direction as that in which a flexible conductor conveying the current from the anode to the cathode would move. They are also deflected by an electrical field in which the lines of force are perpendicular to the direction of the rays, the direction of the deflection being opposite to that in which the lines of force run, *i.e.* in the same direction as a negatively charged body would move. Hence it has been concluded that the cathode rays are formed by negatively charged particles which are projected with great velocity from the cathode. The mass of each of these particles is only about one-thousandth of that of the hydrogen atom, while in a gas at a very low pressure they move with a velocity of about $3{\cdot}6 \times 10^6$ centimetres per second or 80,700 miles per hour. These small charged bodies are called *negative corpuscles* or *electrons*, and the charge carried by each is probably the same as that carried by each hydrogen ion in the electrolysis of a solution.

When the cathode rays strike matter, not only do they induce phosphorescence and produce an increase of temperature, but also in certain cases they give rise to an entirely different kind of radiation. These rays are called Röntgen rays, after their discoverer, or X-rays. They differ from cathode rays in that they pass through glass and many other materials without being greatly absorbed, and they are not deflected by either a magnetic or electrical field. Unlike light they are not refracted when they pass from one medium to another, nor do they show any signs of producing interference phenomena. They probably consist of a disturbance similar to light, but while light is a periodic wave-motion, the X-rays are more in the nature of a single wave or pulse which moves with the velocity of light. These pulses are probably caused by the sudden stopping of the negative corpuscles when they strike a body.

Röntgen rays.

The most striking peculiarity of X-rays is that they are capable of penetrating many substances which are opaque to ordinary light. Thus black paper, wood, and aluminium are transparent, while the more dense

metals, such as lead, are opaque to the rays. The most important practical application of the differences between the transparency of different bodies to these rays is owing to the fact that while flesh is fairly transparent the bones are very much more opaque. Thus if a tube producing Röntgen rays is placed above the hand while a photographic plate is placed below the hand, the rays will pass through the flesh of the hand and will act on the plate. The bones, however, will stop the rays, and hence those parts of the plate within the shadow of the bones will only be slightly affected, and on developing the plate a shadow of the bones of the hand will be obtained. Instead of using a photographic plate, a paper screen which is coated with one of the salts which fluoresces when the Röntgen rays fall on it may be used, when the salt will fluoresce where the rays are transmitted by the flesh, but not where the rays have been absorbed by the bones, and so a dark shadow of the bones will appear on the screen.

Canal rays.

If in a highly exhausted tube a perforated cathode is used, luminous rays are seen to emerge from these holes in a direction away from the anode. These rays are called canal rays, and they are deflected by a magnetic and an electric field in such a way as to show that they consist of positively charged particles. The mass of each particle appears to be about the same as that of the hydrogen atom.

172. Ionisation of Gases.—We have seen how the air in the track of a spark becomes conducting or ionised, and we now proceed to consider this phenomenon in more detail. The fact that the cause of the gas being in the conducting condition can be removed by filtration, and that when the gas is subjected to an electric field the gas loses its conductivity at the same time that a current passes, shows that the conductivity must be due to something of the nature of particles mixed with the gas, and that, whatever this something is, it is charged with electricity so that in an electrical field it moves under the action of the electrical forces. Further, experiment having shown that when a given volume of gas is put into the conducting condition the gas as a whole possesses no charge, it follows that the particles must be some positively and some negatively charged. These electrified particles, to the presence of which the conductivity of the gas may be referred, we may call ions; we shall see, however, that the ions in the case of the conduction of electricity through gases are not the same as the ions in the case of the electrolysis of solutions.

A gas may be ionised in other ways besides by the passage of a spark. Thus a gas is ionised when exposed to Röntgen rays. The passage of electricity through a gas in the conducting state, which for brevity we may call the ionised gas, can be studied by the arrangement

shown in Fig. 306. The ionisation of the gas is caused by the Röntgen ray tube D, which is enclosed in a lead box with a window F. The object of this box is to prevent the escape of the Röntgen rays, except in the direction of the gas under observation, which occupies the space between two insulated metal plates A and B. One of these plates is connected to a quadrant electrometer C (§ 146) and the other to one pole of a battery E. The other pole of the battery is connected to the electrometer. When the tube D is in action the gas between the plates is traversed by the rays and becomes ionised, and a current passes, as indicated by a gradually increasing deflection of the electrometer, the rate at which the electrometer deflection increases being a measure of the current passing through the gas. By measuring the current for different values of the difference of potential between A and B and plotting the current as ordinate and the E.M.F. as abscissa a curve such as that shown in Fig. 307 is obtained. For small values of the E.M.F. the curve is a straight line showing that the current is proportional to the E.M.F., and hence Ohm's law is obeyed. As the E.M.F. is increased the current increases more slowly than would be the case if Ohm's law continued to hold, and a condition is reached when for a very considerable increase of E.M.F. the current remains constant. When, however, the E.M.F. is increased up to a value nearly great enough to cause a spark to pass between the plates A and B there is a very rapid increase of the current. The current corresponding to the flat portion of the curve is called the *saturation current* for the gas under the given conditions.

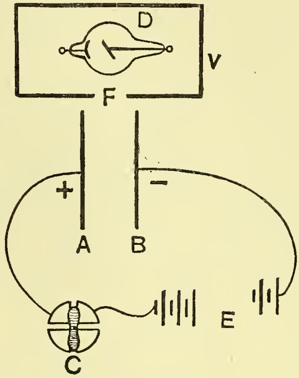


Fig. 306.

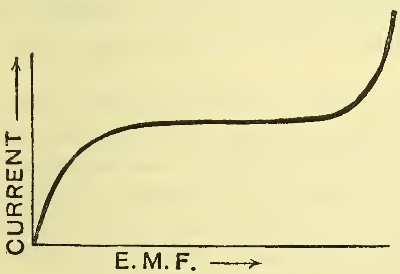


Fig. 307.

The phenomena observed can be explained if we suppose that under the influence of the Röntgen rays ions, that is, positively and negatively charged particles, are produced throughout the gas, the production taking place at a constant rate.

Suppose that in the space between the plates there are q positive and q negative ions produced every second. Then if e is the charge

carried by each ion, and there is a current i passing between the plates, then in a second there will be i/e positive ions driven against the negative electrode B (Fig. 306), and the same number of negative ions driven against the positive electrode A. Now, when a steady state is reached the number of ions removed per second cannot be greater than the number produced by the ionising agent. Hence, if the ionising agent produces q ions of either sign per second the current through the gas cannot be greater than qe, and hence qe is the value of the saturation current. When the E.M.F. between the plates is smaller than that required to produce the saturation current, then since the rate of production of the ions is greater than the rate at which they are removed by the current the supply of ions in the gas will be maintained. The supply will not go on increasing indefinitely, since it is found that there is a limit to the number of ions which a given gas can contain. This limit is probably reached when the rate at which the ions recombine, owing to accidental collisions, is equal to the rate at which the ions are being produced by the ionising agent. Thus as long as the saturation current has not been reached the number of ions between the plates remains constant. The velocity with which the ions travel towards the electrodes depending on the magnitude of the electrical field between the plates, that is, on the E.M.F. applied, the rate at which the ions reach the plates will increase as the E.M.F. is increased. This will go on till the rate of removal of the ions is greater than the rate of production, when, of course, the increased field cannot remove more ions than are produced. When a very high E.M.F. is applied then a new phenomenon makes its appearance, namely, the strong electric field applied to the gas seems to have the power of itself ionising the gas (see p. 521), and hence the supply of ions, and therefore also the maximum current which can pass, is increased. Similar results are obtained with the other ionising agents previously mentioned.

If the ionising agent acts on all the gas between the parallel plates, the ionisation being the same throughout the gas, that is, each cubic centimetre of the gas containing the same number of ions, then the saturation current will increase as the volume of the gas between the plates increases, that is, as the distance between the plates increases. The reason is that the greater the volume of the gas acted upon by the ionising agent the greater the number of ions produced per second, and, therefore, the greater the saturation current. This result shows a very marked difference between the conduction of electricity through gases and that through solids and liquids, for in the case of such conductors the current would decrease as the distance between the plates increases.

The production of the ionisation in a gas at ordinary pressure has been explained by J. J. Thomson in the following manner. An ordinary unionised, *i.e.* uncharged, molecule of the gas is supposed to consist of a number of negatively charged corpuscles clustering round a positive charge, the total negative charge being equal to the positive charge. When the gas is ionised one of the negative corpuscles is torn away from the molecule, leaving the remainder with an excess positive charge. The negative corpuscle then attaches to itself one or more uncharged molecules, the whole constituting the negative ion. The positively charged molecule which has lost a negative corpuscle seems also to attach to itself other uncharged molecules, and the whole constitutes the positive ion. At low pressures the negative ion appears to lose its attached uncharged molecules, so that the negative corpuscle exists alone.

When the ions are in an electrical field they are subject to a force and hence are set into motion. If the field is strong, or if each ion can move for some distance without colliding with other molecules, it may acquire a considerable velocity, and it appears that when a rapidly moving ion strikes an uncharged molecule the impact may be sufficient to ionise this molecule. This production of ions by collision probably takes place just previous to the passage of a spark. There are always a few ions present in air, and if the electric field between two charged bodies is sufficiently strong, these stray ions may be set in motion with sufficient velocity to cause ionisation by collision. The ions at first produced will themselves be set in motion and produce further ions, so that once the field is strong enough to impart the necessary velocity the increase in the number of ions will be rapid, so that the air quickly becomes conducting and a discharge passes. This is also the explanation of the sudden rise of the curve in Fig. 307 for high P.D.'s.

Ionisation is also produced by ultra-violet light, by a red-hot metal, and by flames.

173. Radio-Activity.—Shortly after Röntgen's discovery of the rays which go by his name, Becquerel discovered that uranium, as well as all salts containing this metal, emits rays which are able to pass through black paper and a sheet of thin glass and affect a photographic plate. These rays which are given out by uranium possess the property of discharging both positive and negative electricity, that is, they ionise the gas through which they pass. In fact, the uranium rays behave as regards their electrical and photographical action just as Röntgen rays. Their action is, however, very much more feeble. This power of a body of giving out rays which ionise gases is called *radio-activity*.

Radio-activity is also possessed by thorium and to about the same extent as uranium. By treating large quantities of the mineral pitch-

blende, which contains uranium and thorium and is radio-active, M. and Mme. Curie prepared by chemical means two active substances the radio-activities of which are enormously greater than that of either uranium or thorium. These two substances have been called polonium and radium. They exist in very minute quantities in pitchblende, so that from a ton of the mineral only a decigram or two of an impure salt of radium is obtained. Since then another radio-active substance, called actinium, has been discovered by Debierne.

The radiation given out by radium is very intense. Thus a screen covered with zinc sulphide when brought near a very small quantity (two or three centigrams) of pure radium bromide is brightly lighted up. In the same way a charged electroscope brought near is immediately discharged. One of the most remarkable properties of the salts of radium is that discovered by Curie and Laborde, namely, that they always keep at a temperature of several degrees above that of the surrounding air. Owing to this excess of temperature above that of its surroundings heat must be continuously passing from the radium, and it has been calculated that a gram of radium would emit a quantity of heat of 118 calories per hour. In addition a certain amount of energy is being continually emitted in the form of the rays referred to below. The supply of energy necessary to keep up this continuous flow of energy from radium is derived from the disintegration of the radium atoms, and it has been calculated that the total quantity of energy given out during the disintegration of one gram of radium is about 3×10^9 calories, and is equal to that which would be obtained by burning 500,000 grams of coal. The radium would, however, take about 3000 years to undergo the change, and hence the power available is excessively small.

There are three distinct kinds of rays given out by radio-active bodies. The first of these, called α rays, are very easily absorbed by their passage through matter. Thus the α rays from radium are completely absorbed by a layer of aluminiun ·04 millimetre thick, or a layer of air 7 centimetres thick. The α particles are positively charged helium atoms having a mass 3·8 times that of the hydrogen atom, and are projected from the radium atom with a velocity of about a fifteenth of that of light, but the velocity rapidly decreases as the particles travel through matter. The second type of rays, called β rays, possess all the properties of the cathode rays considered in § 171, and they carry a negative charge. In the case of the β rays projected from radium the average velocity is nearly equal to that of light, and hence is appreciably greater than that of the negative particles in the cathode rays in a vacuum tube, but it is probable that with a voltage of the order of a hundred thousand volts the velocity of the cathode rays would be equal to that of the β rays from radium. The apparent mass of the carrier of the negative

charge in the β rays is about one-thousandth of that of the hydrogen atoms. It is, however, found that the velocity of the β rays produced by different radio-active bodies is different, and that as the velocity increases the apparent mass also increases. Now theory shows that an electrical charge in motion behaves as if it had mass, the apparent mass increasing as the velocity increases. Hence it has been inferred that the β rays consist of negative charges which are not associated with a material nucleus, so that these rays are not matter in the ordinary sense, but disembodied electrical charges which, by virtue of their great velocity, possess the properties of ordinary matter.

The third type of rays, called γ rays, resemble in many ways Röntgen rays, and, like them, consist of a pulse in the ether. This pulse is probably caused by the sudden discharge of each β particle with its high velocity, for it is found the γ rays are only produced when high velocity β rays are present. The γ rays have a very great penetrating power, so that they are able to ionise a gas and thus render it conducting after passing through 30 centimetres of iron.

The three types of rays differ noticeably in respect of the way they behave when passed through a magnetic field which is at right angles to the direction of the rays. Thus the α rays are very slightly deviated, the β rays are strongly deviated, the paths of the particles being circles of comparatively small radius, while the γ rays are apparently undeviated by the magnetic field.

A sheet of aluminium ·004 centimetre thick or a sheet of ordinary writing-paper is sufficient to completely absorb the α rays. The greater part of the β rays can be absorbed by a screen of aluminium having a thickness of ·5 centimetre. Most of the photographic action of the rays given by radium is due to the β rays. The α rays are, however, most effective in producing ionisation in the neighbourhood of a radium salt.

The radio-active bodies, radium, thorium, and actinium, are continually emitting into the surrounding space a material emanation which has the property of a radio-active gas. Uranium and polonium do not, however, give such an emanation. The emanation is able to diffuse through gases and porous substances exactly like a gas, and can be separated from the gas with which it is mixed by the action of cold. Thus when a stream of air which has been passed over radium, so as to become charged with the emanation, is passed through a tube surrounded by liquid air, the emanation condenses in the tube. Passage of the emanation through a platinum tube heated to bright redness has, however, no effect on the emanation.

When the emanation is kept it is found to gradually transform into a gaseous product which exhibits the spectrum of helium and a non-gaseous product which is deposited on the walls of the containing vessel

and is called the active deposit. It is to this deposit that is due the temporary or "excited" radio-active properties acquired by ordinary matter which has been in the neighbourhood of such bodies as radium and thorium. This "excited" radio-activity gradually disappears owing to the transformations which take place in the active deposit.

The phenomena of radio-activity are explained if we assume that the atoms of radio-active elements are gradually and spontaneously undergoing an atomic change. This theory was first put forward by Rutherford and Soddy to explain the results of their experiments on thorium. These observers found that the radio-activity of thorium is due to the presence of a substance which they call thorium-X, and which can be separated from the rest of the thorium by chemical means. After the removal of the thorium-X the thorium is left without activity, the activity being found in the thorium-X. After a time, however, the thorium regains its activity while the thorium-X loses its activity, the rate of gain of the one being equal to the rate of loss by the other. To account for these effects it is supposed that a certain very small proportion of the thorium atoms are continually breaking down, and in the process produce thorium-X. This thorium-X then changes into the emanation and the substance which constitutes the α rays.

A more detailed study of the rates at which thorium-X and the emanation lose their activity and of the character of the rays emitted has shown that the changes are more complex than was at first supposed. As far as is at present known, the changes that take place in thorium are as follows: the thorium atom breaks down, resulting in the expulsion of a helium atom, *i.e.* the production of an α ray, and leaving an atom of a substance called mesothorium 1, which, owing to some atomic change, unaccompanied by the production of any rays, changes into mesothorium 2, and this atom, by the loss of a β particle, which at the same time causes the production of a γ ray, changes into radiothorium. Radiothorium, by the loss of an α particle, changes into thorium-X. Thorium-X, by the loss of an α particle and a β particle, becomes the emanation, while by a further loss of an α particle this changes into the active deposit. This active deposit is not, however, unchangeable. The first product forming the deposit is called thorium-A, and this, by the loss of a β particle, changes into thorium-B. Thorium-B loses an α particle, and changes into thorium-C, while this atom loses another α particle, and becomes thorium-D. Thorium-D loses a β particle, the loss being accompanied by the production of a γ ray, but the final result of all these atomic changes is still unknown, the reason being that while by means of the α, β, and γ rays produced during a change, the fact that such a change is taking place can be determined, yet the number of atoms at any time undergoing the change is so excessively small that no

chemical method is capable of allowing of their detection. It has, however, been suggested that the final product in the case of thorium is bismuth, and that in the case of uranium it is lead. The proportion of atoms which at any moment are taking part in any of the above changes is very different for the different steps. Thus it takes 3×10^{10} years for half the atoms in a given mass of thorium to change into mesothorium 1. Half the mesothorium 1 atoms will, however, change into mesothorium 2 in 5·5 years, while half these will change into radiothorium in 6·2 hours. The times for each of the changes to be half complete, called the transformation period of the change, are given in the following table.

A similar series of changes takes place in the case of the other radioactive substances, and these can be studied in the above-mentioned table. Attention may, however, be drawn to the fact that uranium is an ancestor of radium, and hence the fact that radium is always found in minerals which contain uranium is at once explained.

TRANSFORMATIONS OF THE RADIO-ACTIVE SUBSTANCES.

Substance.	Transformation Period.	Nature of Rays emitted during Transformation.	Substance.	Transformation Period.	Nature of Rays emitted during Transformation.
Thorium ↓	3×10^{10} years	α	Radium-A ↓	3 minutes	α
Mesothorium 1 ↓	5·5 years	...	Radium-B ↓	27 minutes	β
Mesothorium 2 ↓	6·2 hours	β, γ	Radium-C (probably complex) ↓	19·5 minutes	α, β, γ
Radiothorium ↓	2 years	α	Radium-D ↓	15 years	β
Thorium-X ↓	3·7 days	α, β	Radium-E (probably complex) ↓	5 days	β
Emanation ↓	53 seconds	α			
Thorium-A ↓	0·14 second	α	Polonium	140 days	α
Thorium-B ↓	10·6 hours	β	Actinium ↓	Not known	...
Thorium-C ↓	55 minutes	α	Radioactinium ↓	19·5 days	α, β
Thorium-D	3 minutes	β, γ	Actinium-X ↓	10·5 days	α
Uranium ↓	6×10^{9} years	α	Emanation ↓	4 seconds	α
Uranium-X ↓	23 days	β, γ	Actinium-A ↓	0·002 second	α
Ionium ↓	10^{4} years (?)	α	Actinium-B ↓	36 minutes	β
Radium ↓	2000 years	α, β	Actinium-C ↓	2·1 minutes	α
Emanation ↓	3·7 days	α	Actinium-D	4·7 minutes	β, γ

J. J. Thomson has shown that if we suppose that an atom consisting of a number of negatively charged electrons arranged in a ring and rotating with uniform speed within a sphere of positive electrification it would gradually become unstable, owing to the radiation of energy, and either break up with the expulsion of an electron or form a new arrangement. On this theory of atomic structure the changes which take place during the different radio-active transformations can be explained fairly well.

CHAPTER X

ELECTRIC OSCILLATIONS AND WAVES

174. Electrical Oscillations.—If a Leyden jar A (Fig. 308) has a discharge circuit arranged as shown, where C is a coil of wire of fairly low resistance but great self-induction, and D is a spark-gap formed by two well-polished knobs, and the inside and outside coatings are connected to an electrical machine, so as to charge the jar, the following phenomena occur. Suppose the inside coating to be charged positively, then as the P.D. between the coatings is increased, an increasing electrostatic field is produced between the knobs D. As a result the air will finally break down, and a spark will pass. Now the passage of the spark will cause, or at any rate be accompanied by, the ionisation of the air between the knobs, so that the resistance of the spark-gap becomes quite low. As a result the jar is able to discharge through the circuit, and this discharge forms a current passing through the coil C. The current in C will produce a magnetic field, and owing to this field, *i.e.* to the self-induction of the coil, the current will continue to flow even after the jar is completely discharged, causing the jar to charge up in the reverse direction, so that the outside coating is positive. The jar will then begin discharging again, and the current in the circuit will now be in the reverse direction, and will again persist after the jar has lost its charge. Hence the discharge is of an oscillatory nature, the current in the coil being an alternating current. Owing to the heat developed in the coil by the passage of the current, and to other causes which we shall consider in a moment, the amount of electrical energy possessed by the system gradually decreases, and hence the oscillations gradually die out. By examining the spark which passes across the gap in a rotating manner, it can be seen to consist of a number of separate sparks, each of these corresponding to the passage of the discharge in one or other direction.

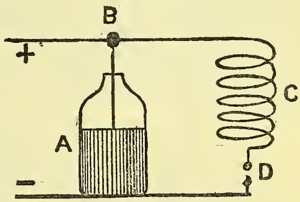


FIG. 308.

An oscillatory system such as that considered above behaves exactly like a spring clamped at one end which is deflected to one side and released, when it will execute vibrations. Now just as the spring will

send out sound-waves through the air, so will the oscillating circuit send out waves. The production of these electro-magnetic waves was predicted by Maxwell, who showed that they had all the properties of light-waves, the only difference being that the wave-length is very much greater than that of light-waves. The existence of these waves was first demonstrated by Hertz, hence they are generally known as Hertzian waves. Hertz obtained stationary vibrations (§ 92) by reflecting electro-magnetic waves from a plane sheet of metal, and from the distance between the nodes he calculated the wave-length. Kelvin having shown how the period of the oscillations of the system which was producing the waves could be calculated, Hertz was able from the period and wave-length to calculate the velocity, and found that this was the same as that of light, a result which Maxwell had predicted many years before.

Electro-magnetic waves.

If two metal rods, AB and CD (Fig. 309), with their rounded ends close together to form a spark-gap are connected to the terminals of an induction coil E, whenever the primary circuit of the coil is broken, the induced E.M.F. in the secondary is sufficient to charge up the wires and cause a spark to occur. After the passage of each spark, the system AB, CD will be set into electrical oscillation, a current passing backwards and forwards across the gap, while each of the ends A and D will be alternately charged positively and negatively. At the instant when A is charged positively and B negatively, an electrical field will exist at the point P, the direction of the field being that of the arrow. As the positive charge at A decreases, the field will decrease in strength till when the end A becomes negatively charged the direction of the field at P will be reversed. Thus the oscillations in AB, CD are accompanied by an alternating electric field at P, the direction of the field being parallel to the wires. When the positive charge at A is decreasing, a current will be passing in the direction from B to C, and this will produce a magnetic field at P, the direction of which is perpendicular to the plane of the paper and upwards. When the positive charge at D is decreasing, the direction of the magnetic field at P will be reversed. Thus the oscillations are accompanied by an alternating magnetic field at P, this field being at right angles to the electric field. Since the current passing at BC is zero whenever the charge at A is at its maximum positive or negative value, it follows that the magnetic field at P is a maximum when the electric field is zero and *vice versâ*. Now we have spoken of the field at P as being a maximum when the charge at A is a maximum, or the current at

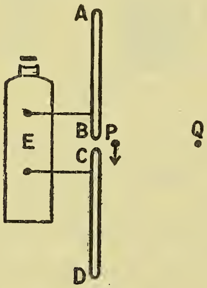


FIG. 309.

BC is a maximum as the case may be. This, however, is only true if P is near BC. If we consider a point Q at a little distance from BC, the maximum of the field strength will occur a little later than that at P, the lag increasing as we go further from BC. This is caused by the fact that it takes time for the field, either electric or magnetic, to spread out. If the oscillations were very slow, such as would be produced if an ordinary alternating current was passed from A to B, the energy which was stored up in the field when the current was a maximum would all return to the conductor when the current was zero. With very rapid alternations, such as we are dealing with in the case of electrical oscillations, a considerable part of the energy of the field does not return to the system, but travels away; in other words, the combined alternating electric and magnetic fields travel away from the oscillating system, and constitute the electromagnetic waves. These waves move in a direction at right angles to AD, and hence the directions of both the electric and magnetic fields are at right angles to the direction of propagation of the waves, that is, they are

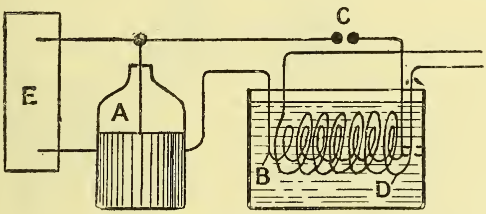


FIG. 310.

in the wave-front, and hence they constitute a transverse wave. By examining the reflection of electro-magnetic waves at the polarising angle, it has been shown that what is called the plane of polarisation in light is the plane in which the magnetic field acts, the electric field, therefore, being at right angles to the plane of polarisation.

175. Tesla Coil. Wireless-Telegraphy.—The oscillations produced by the discharge of a Leyden jar have been utilised to supply alternating currents of extremely high frequency in the Tesla coil. The inside and outside coatings of a Leyden jar, A (Fig. 310), are connected to a coil of wire B, a spark-gap C being included in the circuit. A secondary coil D surrounds the coil B, both coils being immersed in oil so as to prevent a spark from passing from one coil to the other. The jar is connected to an induction coil, and each time a spark passes at D oscillations are set up in the primary coil which by induction produce an oscillatory current in a circuit connected to the secondary coil D, this oscillatory current being of a very high frequency.

A practical application of Hertz waves consists of their employment

in the so-called wireless-telegraphy for the transmission of signals from one place to another without the necessity of a metallic wire connecting the two stations, such as is used in the ordinary telegraph. The inception of this application of electro-magnetic waves, and much of the development in the practical application of the method, is due to Marconi. A diagrammatic representation of Marconi's arrangement for the transmission of wireless messages is shown in Fig. 311. The left-hand side shows the trans-

Wireless-telegraphy.

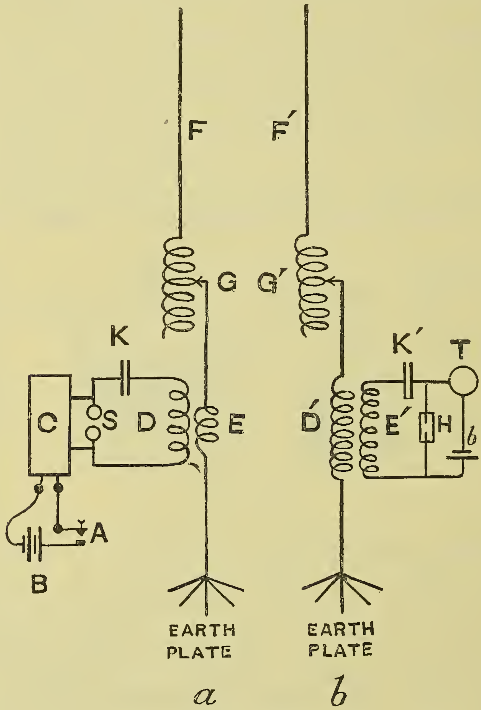


Fig. 311.

mitting arrangement, which consists of an induction coil C, connected to a battery B, and key A. The secondary of the coil is connected to the spark-gap S, which forms part of an oscillation circuit containing a condenser K, and a coil of a few turns D. Near this coil D is placed another coil, E, which is connected at one end to an earth plate, consisting of a series of radiating wires buried in the earth, and at the other end is joined to a sliding contact on the turns of a coil G. One end of the coil G is connected to a wire, or series of wires, F, supported on a tall mast, forming what is called the antenna or aerial. When the key A is

depressed the contact-breaker on the coil works, and a series of sparks pass at S, while electrical oscillations are set up in the circuit containing D and K. The oscillations in the coil D induce oscillations in the neighbouring coil E, these oscillations taking place throughout the antenna, whence electro-magnetic waves are radiated, and travel away with the velocity of light. The object of the sliding contact on the coil G is to allow of tuning the antenna to resonance with the circuit DSK by altering the number of turns in the coil, and hence its self-induction. This turning is of considerable importance when it is required to produce energetic oscillations in the antenna, such as are required for transmitting signals to a considerable distance. The receiving arrangements are shown on the right of the figure. The antenna F′, which, owing to the incidence of the waves, is set into electrical oscillation, is connected through the coils G′ and D′ to earth, and by means of the sliding contact on G′ the natural period can be adjusted to unison with that of the transmitting station. A coil E′ is placed near to D′, and this coil is connected to a condenser K′ and a receiver L, which may be a coherer (see below) or some similar device. When the oscillations are produced in the receiving antenna, owing to induction, oscillations will also be produced in the circuit E′K′H, and these will cause the coherer to become conducting, so that the battery *b* can send a current, and hence the galvanometer T is deflected. When a coherer is employed as detector, some arrangement is provided to tap the tube and cause it to decohere. A form of detector often employed consists of a pointed piece of graphite touching the face of a crystal of galena, the contact forming part of the circuit. Such a contact allows the passage of a current better in one direction than the other, and hence is often called a rectifier. When it is employed a telephone is often used at T in place of a galvanometer to read the signals.

The coherer is an instrument invented by Branley for the detection of electrical oscillations. An improved form of coherer, designed by Marconi, consists of two amalgamated silver plugs sealed in a glass tube and connected to two wires which pass through the glass. The two plugs are separated by a space of about a millimetre, which is partly filled with filings of an alloy of nickel and silver. The tube is exhausted of air and then sealed up. When such a coherer is included in the circuit of a battery and galvanometer, the filings form what is practically an insulator, so that no current passes. On producing electric oscillations in the neighbourhood of the circuit, however, the tube containing the filings becomes a conductor and the battery is able to drive a current through the circuit, so that the galvanometer is strongly deflected.

CHAPTER XI

THERMO-ELECTRICITY

176. Thermo-Electric Junction.—In 1821, while making experiments on the difference of potential which appears to exist between two different metals when placed in contact, Seebeck noticed that if a circuit is formed which is composed of two wires of different metals joined together at their ends, and if the junctions are at different temperatures, a current will in general be produced in the circuit. Thus if two copper wires, which are connected to the terminals of a galvanometer, are connected at their other ends to a piece of iron wire, and one of the junctions of the copper and iron is heated, a current will be indicated by the deflection of the galvanometer. The direction of the current will be from the hot to the cold junction in the iron. This current is said to be a thermo-electric current.

If, while the cold junction is kept at a constant temperature, the temperature of the hot junction is gradually raised, the current in the circuit will gradually increase up to a certain point, this temperature being called the neutral point for the two given metals. If the temperature of the hot junction is raised above the neutral point the current in the circuit will decrease, till, when the temperature of the hot junction is as much above the neutral point as that of the cold junction is below, there will be no current in the circuit; while if the temperature of the hot junction is yet further raised, the direction of the current will be reversed.

It is possible to arrange the metals in a series such that if wires of any two of them are joined together to form a circuit, and one of the junctions is heated, the thermo-electric current will in the first metal on the list go from the hot junction to the cold, it being supposed that the mean temperature of the hot and cold junctions has some given value. The following is such a thermo-electric series for a mean temperature of about 50° C.: antimony, iron, zinc, silver, tin, copper, bismuth. Of course the order of the metals will vary with the temperature, for the neutral temperature for some of the combinations is quite low, and the neutral points for the different combinations vary very much.

It is found that if we have three metals A, B, and C, and if the E.M.F. developed in a circuit consisting of A and B when the tempera-

tures of the junctions are t_1 and t_2 is E_{AB}, then if a circuit be formed of the three metals, and the junctions of A and C and B and C are kept at the temperature t_1 and that of A and B at t_2, the E.M.F. developed will still be E. Hence the interposition of the metal C, so long as the temperature of the two junctions with this metal are the same, has no influence on the E.M.F. developed by the other two metals. This result can be extended to any number of additional metals.

If one of the junctions of a thermo-couple is kept at a constant temperature, say 0° C., and the temperature of the other is varied, the E.M.F. developed in the circuit is not in general proportional to the difference in temperature between the junctions. For pure metals the curve showing the relation between E.M.F. and temperature difference is approximately a parabola with its axis parallel to the axis of E.M.F., the temperature corresponding to the vertex of the parabola being the neutral temperature.

The change in E.M.F. per degree difference in temperature of the hot junction, the cold junction being at 0°, is called the *thermo-electric power* of the two metals, and a diagram showing the thermo-electric powers of the metals when combined with lead at different temperatures is called a thermo-electric diagram.

Thermo-couples are often employed for measuring temperatures, in which case one junction is generally immersed in melting ice and the couple is calibrated by noting the E.M.F. developed when the other junction is raised to known temperatures. For comparatively low temperatures, such as are involved in measurements of radiant heat, couples consisting of pure iron and nickel or copper and an iron alloy called constantan are generally employed. These couples give an E.M.F. of about 30 microvolts per degree centigrade. When used to measure radiant heat, one end of each of the wires is soldered to the back of a very thin disc of blackened copper, the other ends of the wires being soldered to copper wires which are led off to a delicate galvanometer. The junctions with the copper are kept at a constant temperature, and the temperature of the other junction is deduced from the galvanometer deflection.

For measuring high temperatures, such as that of a furnace, a thermo-couple consisting of pure platinum and an alloy of platinum containing 10 per cent rhodium is generally employed.

177. The Peltier and Thomson Effects.—In a thermo-electric circuit of which the junctions are at different temperatures, there is a current flowing; and we have seen that the passage of a current through a conductor involves the expenditure of some energy, which appears as heat according to Joule's law. The question now arises in what manner the energy necessary for the maintenance of the current in the

thermo-electric circuit is supplied. This question is answered by a discovery made in 1834 by Peltier, who found that when an electric current is passed through a thermo-electric junction, *i.e.* a junction of two different metals, there will be either a development of heat at the junction or an absorption, according to the direction in which the current is passed.

The Peltier effect differs from the Joule heating already considered, in that while the Joule heating is proportional to the square of the current, and is independent of the direction of the current, the heat developed at a junction of two metals is proportional to the first power of the current, and depends on the direction of the current.

Now let us consider a thermo-electric circuit composed of iron and copper, and suppose that one of the junctions is immersed in a mixture of ice and water, while the other junction is placed in a beaker of water at a temperature t. Then we know that a current will flow, the direction of which at the hot junction will be from the copper to the iron, while at the cold junction it will be flowing from the iron to the copper. Now when a current flows from copper to iron there is, according to Peltier's observation, an absorption of heat, while when a current flows from iron to copper there is a liberation of heat. Hence in our example there will be an absorption of heat at the hot junction, which will be supplied from the heat of the hot water, while there will be a liberation of heat at the cold junction, which will melt some of the ice. We thus see that the production of a thermo-electric current is accompanied by a transfer of heat from the source which is used to maintain the temperature of the hot junction to the refrigerator used to cool the cold junction. If then the heat given up to the refrigerator, or the condenser, as we may call it from the analogy with the thermal engines considered in the sections on thermo-dynamics, is less than the quantity absorbed from the source by the amount of heat developed in the circuit according to Joule's law, the maintenance of the current is at once accounted for. We are thus led to look upon a thermo-electric circuit as an ordinary heat engine in which a certain quantity of heat is taken in at a given temperature; some of the heat used up by the engine in this case is first converted into electric energy, and then reconverted into heat, according to Joule's law, but might in part at least be converted into mechanical work, while the remainder of the heat taken from the source is given out to a condenser which must be at a lower temperature than the source.

There is another reversible thermal effect when a current is passed along a conductor of which the temperature varies from point to point, which was discovered by Lord Kelvin (William Thomson), and is called the Thomson effect. He found that in addition to the heat developed

according to Joule's law, which is proportional to the square of the current, there is a slight heat development which is proportional to the current when the *temperature of the conductor varies from point to point.* Whether the heat is absorbed or evolved depends on the relation between the direction of the current and the temperature gradient, and on the nature of the metal. Thus in the case of copper, heat is absorbed when the current flows from the cold to the hot part of the wire, while for iron the reverse is the case. If the current flows from the hot part to the cold, heat is evolved in the case of copper and absorbed in the case of iron.

QUESTIONS AND EXAMPLES

Except when otherwise mentioned, the following values for some constants have been employed when working out the answers:—

$g = 32$ ft./sec.^{2} or 980 cm./sec.^{2}.
$J = 4{\cdot}2 \times 10^7$ dynes per calorie or 778 ft.-lbs. per B.Th.U.
Density of mercury, 13·6.
Weight of a cubic foot of water, 62·42 lbs.

BOOK I

CHAPTERS I, II, AND III

1. Explain what is meant by a velocity curve and show how such a curve may be used to determine the space traversed between two given instants.

Draw the velocity curve, and by measuring an area deduce the total space traversed in 9 seconds in the case of a body which moves as indicated in the following table:—

Time	0	1	2	3	4	5	6	7	8	9	seconds.
Speed	0	17·4	34·2	50·0	64·3	76·6	86·6	94·0	98·5	100·0	cm./sec.

2. If a ship can sail at an angle of 45° to the wind, with a speed of 5 knots, how long will it take to go 20 nautical miles to windward if the nearest it can sail to the wind is 45°?

3. Show that when a particle moves with uniform speed in a circle it has an acceleration towards the centre, and find an expression for this acceleration.

4. If the earth has a radius of 4000 miles, what is the acceleration of a point on the equator due to the earth's rotation?

5. Find the constant force which must act on a mass of 50 grams to increase its velocity by 100 cm./sec. in 20 seconds.

6. For how long must a constant force of 100 dynes act on a body originally at rest, of which the mass is a kilogram, to produce a speed of 50 cm./sec.?

7. Two men, A and B, carry a weight slung from a pole; if the maximum weight A can carry is 100 lbs., and the maximum weight B can carry is 150 lbs., and the length of the pole is 10 ft., show at what point the largest load which can be carried must be supported.

8. A uniform beam AB, which weighs 100 lbs., is 16 ft. long, and rests on two supports, one at 4 ft. from A and the other at 5 ft. from B. (*a*) Calculate the pressure exerted on the supports when a boy weighing 60 lbs. stands on the end A. (*b*) What will be the maximum weight which can be placed on the end of the beam without the beam tilting up?

9. Explain how the units of force are defined in the metric and British systems respectively.

If a kilogram is equal to 2·2 lbs. and an inch is equal to 2·54 cm., find the ratio of the units of force in the two systems.

10. Prove that the kinetic energy of a body of mass m moving with a velocity v is $\frac{1}{2}mv^2$.

11. Define the units *horse-power*, *erg*, *watt*.

Find the horse-power required to propel a ship at 20 miles an hour, the resistance to the ship's motion at that speed being equal to the weight of 60 tons.

12. If to maintain a speed of 10 miles per hour along a level road a cyclist has to exert a power of a tenth of a horse-power, what is the resistance to motion which he experiences?

13. What amount of work against gravity is performed when a man pumps 1000 gallons of water to a height of 40 ft. in 10 minutes, and what horse-power is exerted if we suppose the pump, &c., to be without friction?

14. Find the horse-power required to raise 1000 lbs. of water per minute through a height of 20 ft., the water leaving the top of the pipe with a horizontal speed of 16 ft./sec.

15. A ladder 25 ft. long rests against a wall and makes an angle of 30° with the vertical. How much work is performed against gravity by a man weighing 160 lbs. when he ascends the ladder?

16. A mass of 3 lbs. and a mass of 5 lbs. are connected by a string which passes over a pulley. Neglecting friction and the mass of the pulley, find the kinetic energy of each mass and the tension in the string when the system has moved through 16 ft. from rest.

CHAPTERS IV, V, AND VI

1. Explain what is meant by the terms angular velocity, angular acceleration, moment of inertia, angular momentum, and torque, and give any relations which exist between these quantities.

2. A cylindrical rod is a metre long and consists of 80 cm. of wood of which the mass is 200 grams, and 20 cm. of metal of which the mass is 500 grams. Find the position of the centre of gravity.

3. Find the position of the centre of gravity of a thin plate in the form of a square, each side 10 cm., from which a triangular portion included between one side and the two diagonals has been removed.

4. Find the force which must act on a particle that it may move with a uniform speed in a circular path.

What would be the length of the day if the speed of rotation of the earth were to increase so that a body at the equator would just fly off? The radius of the earth may be taken as 4000 miles and g at the equator as 32 ft./sec.^{2}.

5. A flywheel weighs a ton, and the whole weight may be supposed concentrated at the rim, which has a mean radius of 5 feet. What is the energy stored in the flywheel when it is making 240 revolutions per minute?

6. If the above flywheel when left to itself is acted upon by a uniform torque resisting its motion and comes to rest in 2 minutes, what is the value of the torque?

7. Explain what is meant by friction, and define the terms *coefficient of friction*, *angle of friction*. Show that when a body is just ready to slip down an inclined plane the tangent of the angle of inclination of the plane is equal to the coefficient of friction.

8. In what way do weight, inertia, and friction respectively affect the force which has to be used to propel a vehicle?

Why is it more important to avoid steep inclines on a railway than on a road?

9. Explain what is meant by the efficiency of a machine, and prove that as the friction of the different parts is reduced the efficiency increases but can never be greater than unity.

10. A thin rope is wound round the flywheel of an engine of which the diameter is 10 ft., and when the speed of the engine is 300 revolutions per minute the difference in the tensions at the two ends of the rope is 40 lbs. What is the horse-power of the engine?

11. Show that with a Barton tackle the load supported by the tackle will not of itself be able to move down unless the efficiency is greater than 0·5.

12. If a motor-car, when the engine is disconnected, is able to coast down a slope of 1 in 20 at a uniform speed of 20 miles per hour, calculate the resistance to motion at this speed in pounds per ton of the car's weight.
13. What power must the engine develop in the case of the car considered in the above question to propel the car at 20 miles per hour (a) along the level and (b) *up* the slope, if the car weighs a ton?
14. What are the most important qualities of a good balance, and how are they secured? How would you determine whether the arms of a balance were of equal length, and how would you eliminate error due to such an inequality?
15. Two weights of the same nominal value are placed in the two pans of a balance. If the pointer comes to rest at a point on one side of the zero, how would you proceed to determine whether the error is in the weights or in the balance?

CHAPTERS VII AND VIII

1. Explain what is meant by the terms strain, stress, simple shear, and coefficient of rigidity.
   Opposite vertical faces of a cube of a given material are attached to two plates. One plate is held fixed, and a weight of 10 kilograms is suspended from the other, and causes this plate to descend through a distance of 2 mm. If each edge of the cube measures 10 cm., calculate the coefficient of rigidity of the material.
2. A cubical block of a solid is subjected to a uniform compression perpendicular to two opposite faces and an equal uniform tension perpendicular to another pair of faces. Show that the resultant strain is a shear.
3. Explain what is meant by Boyle's law.
   If a barometer has a little air in the space at the top of the mercury column, and, as a result, reads 28·5 in. when it ought to read 29 in., and reads 29 in. when it ought to read 30 in., what is the real value of the atmospheric pressure when this barometer reads 29·25 in.?
4. Define a coefficient of elasticity. Show that the coefficient of volume elasticity of a perfect gas is numerically equal to the pressure when the temperature is constant.
5. Describe some arrangement for the measurement of very small gaseous pressures.
6. Explain the action of the syphon, and calculate the greatest depth of a vessel from which a liquid of which the density is 2 can be removed by means of a syphon when the barometer stands at 30 in. (Density of mercury = 13·6.)
7. The volume of a steel bottle is half a cubic foot, and air is pumped into the bottle till the pressure rises from atmospheric to 50 atmospheres above atmospheric. What volume of air, measured at atmospheric pressure, has been forced in?
8. Describe and explain the mode of action of some form of mercury pump.
9. If the volume of the receiver of an air pump is 2 litres and the pump has a stroke of 20 cm., the cross-section of the piston being 3 sq. cm., what will be the pressure in the receiver after three strokes if the initial pressure corresponds to 76 cm. of mercury and the clearances can be neglected?
10. Describe experiments which have been made on the relation between the volume and pressure of a gas for great ranges of pressure, and give the general character of results obtained.

CHAPTERS IX, X, AND XI

1. Give a proof of the principle of Archimedes.
   A Nicholson's hydrometer requires a weight of 10 grams to sink it to its mark. When a body is placed in the upper pan it requires 7 grams to sink the hydrometer to the mark, while when the body is in the lower pan 9 grams is sufficient. What is the volume and density of the body?
2. A lock gate is 20 ft. wide; what will be the total pressure on the gate if the level of the water is 10 ft. higher on one side than on the other? If the difference in level were halved would the total pressure be halved?

3. If the density of ice is ·917 and that of sea-water 1·025, what fraction of the volume of an iceberg is above water?
4. A uniform thin cylinder consisting of a length of 20 in. of wood of which the specific gravity is ·75, and half an inch of metal of which the specific gravity is 6, is floated upright in water; how much of the cylinder will be above the surface?
5. In the above example will the cylinder when floating upright be in stable equilibrium, and where approximately will the metacentre be situated?
6. Show how to find the resultant pressure of a heavy liquid on an immersed body.
   A block of wood 6 ft. square and 1 ft. thick is floating in sea-water of which the specific gravity is 1·025. If the specific gravity of the wood is 0·6, what weight of iron can be placed on the block without it being entirely immersed?
7. In what proportion should two liquids, A and B, be mixed so that the mixture may have a density of 1·4, the density of A being ·9 and that of B 1·6?
8. Find an expression for the velocity of efflux of water through a circular hole in the bottom of a vessel, the head of water in the vessel being kept constant.
9. What is meant by the surface tension of a liquid? Calculate the height to which a liquid will rise in a capillary tube. How would your calculation be affected by a want of uniformity in the bore of the tube?
10. Show that the energy of a soap film is equal to $2T$ per sq. cm. where T is the surface tension of the liquid. How much work against surface tension has to be performed to blow a soap-bubble 10 cm. in diameter if the surface tension of soap solution is 49 dynes per cm.
11. Two vertical parallel glass plates, 0·2 of a millimetre apart, are immersed in water for which the surface tension is 80 dynes per cm., the angle of contact being 0°. Find the height to which the water will rise at some distance from the edges of the plates.
12. Explain what is meant by Young's modulus.
    If Young's modulus for a steel wire is 2×10^{12} dynes/cm.^{2}, what weight must be suspended from a wire 2 metres long and 1 mm. in diameter to stretch it by 1 mm.? ($g = 980$ cm./sec.^{2}.)
13. Explain how Young's modulus for a material may be deduced by observing the extension of a wire. Draw the stress-strain diagram for this case, and show how the work performed in straining the wire can be obtained from such a diagram.
14. A uniform tube of length L, filled with air at atmospheric pressure, of which the upper end is closed, is lowered mouth downwards in water. Show that if H is the height of a water barometer, and the water rises in the tube through a distance x, then the depth D to which the tube has been lowered is given by $D = \dfrac{Hx}{L - x}$.
15. An open trough of which the vertical cross-section is a square, each side 50 cm., is filled with water. One of the vertical ends of the trough is hinged at its upper edge. Find the least force which applied to this end will keep it closed.

BOOK II

CHAPTER I

1. A glass rod when measured with a zinc scale, both being at 20° C., appears to be a metre long. If the scale is correct at 0° C., what is the true length of the glass rod at 0°? The coefficient of linear expansion of glass is 0·000008, and that of zinc 0·00003.
2. A barometer with a glass scale reads 760 mm. at 17° C. Find the correction required to reduce the readings to 0° C., if the apparent coefficient of expansion of mercury in glass is ·0001558, and the coefficient of linear expansion of glass is ·0000087.
3. Describe an absolute method of determining the expansion of mercury with rise of temperature. Knowing the expansion of mercury, describe exactly how you would find the expansion of another liquid between 0° and 100° C.

4. How has the temperature of maximum density of water been determined?

A glass bulb is loaded so that it just sinks in water at 0° C. What will happen to it as the water is gradually raised in temperature to 20° C.?

5. A piece of metal for which the coefficient of linear expansion is a, weighs w grams in air, w_1 grams when suspended in a liquid at 0°, and w_2 grams in the liquid at t°. Show that the coefficient of expansion of the liquid is

$$\frac{1}{w-w_2}\left\{\frac{w_2-w_1}{t}+3a(w-w_1)\right\}.$$

6. Explain what is meant by the coefficient of apparent expansion of a liquid, and prove the relation between this quantity and the true coefficients of cubical expansion of the liquid and envelope.

7. The volume of a flask is 1000 c.c., and it is filled with air at 15° C. and atmospheric pressure. The flask is heated to 100° C., and it is found that 290 c.c. of air, measured at the original temperature and pressure, must be allowed to escape before the pressure in the flask falls to its original value. Calculate the coefficient of expansion of air, neglecting the expansion of the flask.

8. The density of mercury at 10° is 13·57, and that at 100° is 13·35. What is the mean coefficient of expansion between these two temperatures?

9. A glass bulb when empty weighed 4·023 grams; when filled with water at 22° it weighed 13·828 grams, and when filled with water at 75° it weighed 13·601 grams. If the mean coefficient of cubical expansion of the glass is ·0000184, calculate the mean coefficient of expansion of the water.

10. If the density of air at 0° and 76 cm. pressure is ·001293, what is the density at 30° and at a pressure of 72 cm.?

11. From the numbers given in the previous question calculate the value for air of the constant R in the equation $pv=RT$.

12. Describe the constant volume air thermometer, and explain how you would use it to determine the temperature in the centigrade scale corresponding to the absolute zero on the gas thermometer.

CHAPTERS II AND III

1. Describe how you would determine the latent heat of fusion of ice.

If 50 grams of pounded ice at 0° C. are mixed with 100 grams of water at 20°, what will be the final state of the mixture? The thermal capacity of the calorimeter may be neglected.

2. Explain what is meant by the term latent heat.

If the temperature of a kilogram of ice, originally at −10° C., is gradually raised to 110° C., the pressure remaining one atmosphere throughout, describe the changes in volume and state which occur, and give approximately the quantities of heat which are absorbed by the substance in the different stages.

3. A piece of copper, specific heat 0·1 and mass = 150 grams, is cooled to −150° C., and then plunged into a calorimeter containing 200 grams of water at 10° C. What will be the result when temperature equilibrium has been reached? The water value of the calorimeter is to be neglected.

4. When 10 grams of steam at 200° C. and 760 mm. are condensed in a copper calorimeter weighing 400 grams and containing 1000 grams of water at 15° C., the temperature rises to 21·5° C. If the latent heat of steam at 100° C. is 540 calories per gram, and specific heat of copper ·095, find the specific heat of steam.

5. What is meant by the statement that the maximum pressure of water vapour at 0° is 0·46 cm. of mercury? Explain what happens when water at 10° is placed under the receiver of an air-pump and the air is exhausted.

6. A copper calorimeter weighs empty 8 grams, and when full of water at 15° C. weighs 308 grams. Steam is passed into the water till the weight rises to 310 grams. If the specific heat of copper is 0·1, find the final temperature.

7. Describe how Joly has directly determined the specific heat of a gas at constant volume.

8. Describe Bunsen's ice calorimeter, and enumerate the operations and observations which are necessary when calibrating the instrument and using it to determine the specific heat of a metal.

9. Describe Bunsen's ice calorimeter, and explain how you would use the instrument to determine the specific heat of a solid.

If 1 gram of water expands by ·0918 c.c. on freezing, and evolves 80 calories, calculate the specific heat of a substance, 5 grams of which when introduced into a Bunsen's ice calorimeter at a temperature of 100° C. produces a contraction of 700 cubic mm.

10. Describe some method by which the specific heat of a gas at constant pressure can be determined.

11. How can the vapour pressure of water be determined at temperatures higher than 100°.

If you were given a curve showing the vapour pressure of water, explain how the pressure-gauge of a steam-boiler might be replaced by a thermometer.

12. Explain what is meant by the steam line, the hoar-frost line, and the ice line, and what is indicated at the point where they meet.

13. Why do two thermometers, one with its bulb dry and the other with its bulb moist, generally indicate different temperatures when exposed to the air at the same place?

Under what conditions would the readings of the two thermometers be identical?

14. Explain how the humidity of the air can be obtained by determining the dew-point. What data would you require and what assumptions would you make in obtaining your result?

15. If some of the air of the room is introduced into a bottle which is then corked, what change, if any, would take place (1) in the dew-point and (2) in the relative humidity of the contained air if the temperature of the bottle is raised?

16. Ice, water, and salt are mixed, two and two, in approximately equal proportions, each being originally at 0° C. Which of the three mixtures will be at the highest and which at the lowest temperature? Give reasons for your answer.

17. Explain what is meant by a eutectic, and show that the composition of the eutectic of, say, copper-silver can be deduced from the cooling curves for a number of alloys.

18. Explain what is meant by the calorific power of a fuel.

Given that 1 gram of carbon when burnt to carbon dioxide liberates 7860 calories, and when burnt to carbon monoxide it liberates 2140 calories, what proportion of the energy of a carbonaceous fuel is wasted if the flue gases contain equal volumes of CO_2 and CO.

CHAPTER IV

1. Explain carefully how heat is conveyed through a liquid by convection, and describe some practical application of this phenomenon.

2. How do you explain the fact that if a piece of metal gauze is held above a gas-burner, and the gas is lighted above the gauze, the flame does not go below the gauze?

3. Describe and explain the action of the miner's safety-lamp. When does the lamp become dangerous?

4. Give a careful definition of thermal conductivity.

Estimate the rate at which ice will melt in a wooden box 2 cm. thick, of inside measurements $100 \times 60 \times 60$ cm., assuming that the conductivity of wood is 0·00040 and that the external temperature is 20° C.

5. How much heat will be transmitted through each square metre of an iron boiler plate, 4 mm. thick, in an hour if the opposite faces are kept at 300° and 120° respectively? The conductivity of iron is 0·2.

6. The outside of a glass beaker is kept at 100° C., and it contains a mixture of ice and water which is kept well stirred. If the thickness of the glass is half a millimetre and the area is 200 sq. cm., find the amount of ice melted per minute if the conductivity of glass is ·002 in C.G.S. centigrade units.

7. A boiler is made of iron, 0·6 cm. thick and of conductivity 0·2. A crust 0·6 cm. thick is formed inside of conductivity 0·01. If the outside of the iron is at 300°, and the inside of the crust is at 150°, what quantity of heat is conducted through each square centimetre per second?

CHAPTER V

1. Give an account of an accurate method of determining J.
   A gram of coal in burning gives out 8000 calories. With what velocity must a mass of coal be moving to contain the same amount of mechanical energy as it would give out in burning?
2. With what velocity must a lead bullet strike a target in order that its temperature may rise 100°, supposing half the heat due to the impact to be retained by the bullet? (Specific heat of lead $=0{\cdot}03$.)
3. A paddle turns about a vertical axis in a calorimeter which contains water and is pivoted so that it can turn about a vertical axis. If a couple of 10^7 dynes-cm. has to be applied to prevent its rotation, calculate the rise in temperature of the water while the paddle makes 1000 revolutions, if the water equivalent of the vessel and its contents is 2000.
4. Explain what is meant by the terms specific heat of a gas (1) at constant pressure and (2) at constant volume.
   Explain carefully why one is greater than the other, and what assumptions would enable you to calculate the difference of these specific heats at any pressure if you were given the value of the mechanical equivalent.
5. Find an expression for the work done by a gas when expanding against a constant pressure.
6. Describe Joule and Kelvin's method for determining the work performed against molecular forces when a gas expands, and the way in which the effect observed has been applied to the liquefaction of air.
7. Explain what is meant by (a) an isothermal and (b) an adiabatic curve, and show how the work performed during an isothermal or adiabatic change in the case of a gas can be represented on the p-v diagram.
8. What is the critical point for a gas, and what is the general form of the isothermals for temperatures near the critical temperature?
9. Find an expression for the adiabatic elasticity of a gas.
10. Describe the Carnot cycle, and find an expression for the efficiency of an engine working on this cycle between given temperature limits.
11. Show that the efficiency of an engine working between any two temperatures is a maximum if the cycle on which it works is reversible.
12. What is the efficiency of an engine working on a reversible cycle between the temperature limits 200° C. and 0° C.?
13. Explain how Kelvin's absolute scale of temperature is arrived at.
14. If a mass of saturated air is allowed suddenly to expand a cloud is formed. Explain this phenomenon.

BOOK III

CHAPTERS I AND II

1. Draw the harmonic curves for two S.H.Ms., the periodic time for one being 12 seconds and for the other 3 seconds, the amplitude of the first being 4 times that of the second. Also draw the curve which represents the result of adding these together, the phase at the start being the same when the displacements are zero.
2. A particle executes a S.H.M. Find expressions for the velocity and acceleration in terms of the amplitude, periodic time, and the displacement.
3. Explain what is meant by the phase of a S.H.M.
   Plot a curve showing the result of compounding two S.H.Ms. of equal period and amplitude in directions at right angles, the phases differing by 90° or a quarter period.
4. Find the resultant of two S.H.Ms. in the same direction and having the same period, the phase of one being 90° ahead of the other; the amplitude of the more advanced being twice that of the other.
5. Find an expression for the period of a simple pendulum.
6. Explain what is meant by simple harmonic motion. If a particle moves backwards and forwards through a distance of 20 cm. five times a second with a S.H.M., find its greatest velocity and its greatest acceleration.
7. Find an expression for the velocity with which the bob of a simple pendulum

passes through its position of rest, and show that if the amplitude of the motion is small, the velocity with which the bob passes through its lowest position is proportional to the amplitude; given that $1 - \cos\theta = 2\sin^2\frac{\theta}{2}$.

8. Given that the square of the radius of gyration of a uniform cylindrical rod about an axis through the mid point perpendicular to the length is equal to $l^2/12$, where l is the length, calculate the period of such a rod of length 100 cm., when oscillating about axes perpendicular to the length through points at different distances from one end, and plot a curve showing the connection between the period and the position of the point of support. From this curve determine the minimum period and show that it is equal to that of a simple pendulum of length $l\sqrt{3}/6$.

9. A circular disc of radius 10 cm. is suspended from a point in its circumference so as to be able to turn about a horizontal axis perpendicular to its plane. Given that the square of the radius of gyration of a circle about an axis through the centre perpendicular to its plane is equal to half the square of the radius, find the period of the disc when it executes small oscillations.

10. Distinguish between forced and free vibrations, giving an example of each.

11. Why is it necessary to compensate the pendulum of a clock to allow for the effect of temperature changes. Describe some form of compensated pendulum, and calculate the ratio of the length of the rods if they consist of zinc and iron; given that the coefficient of linear expansion of iron is 0·000012 and that of zinc 0·000030.

12. Explain what is meant by a stationary wave motion, and describe the condition of the surface particles at different points at different times when stationary waves are produced on the surface of water.

13. Describe an experiment for showing the interference fringes produced when two series of waves of the same period are generated at two adjacent points.

14. A simple pendulum of mass m (the mass of the cord can be neglected) is held deflected from the vertical through an angle θ by a horizontal force F applied to the bob. Show that the tension in the cord is given by

$$T = \frac{mg}{\cos\theta} = \frac{F}{\sin\theta}.$$

Further show that if, when the pendulum is at rest in this position, F is removed, the tension changes to the value $T = mg\cos\theta$.

Find F and the two values of T for $m = 4000$ grams, and $\theta = 30°$.

15. Show that there are two parallel axes at different distances from the centre of gravity about which the period of a compound pendulum is the same.

CHAPTER III

1. Upon what properties of a fluid does the velocity of sound in it depend? At what temperature would the velocity of sound in air be twice as great as it is at 0° C?

2. The velocity of longitudinal vibrations along a steel rod of density 7·7 is 4900 metres per second. What is the value of Young's modulus for the steel?

3. Sound-waves travelling against the wind appear to have their direction of propagation deflected upwards. Account for this phenomenon.

4. If the product of the pressure and volume of 1 gram of air is 18·4 joules at 0° C., and if the velocity of sound at 0° C. is 332 metres per second, find the ratio of the two specific heats of air. (1 joule $= 4{\cdot}2 \times 10^7$ ergs.)

5. A source of sound is moving with velocity u towards a hearer. Find the alteration in pitch of the note heard.

6. A tuning-fork is excited and moved while sounding towards a wall. Explain the beats which are heard.

If the fork makes 500 vibrations per second, and when it is moved at a uniform speed towards a wall 2 beats per second are produced, what is the speed? (Velocity of sound = 332 metres per second.)

7. A train of sound-waves is reflected by a fixed plane at right angles to the direction of propagation. Describe the nature of the resulting motion, and explain how it may be investigated experimentally.

8. Explain the phenomenon of the "beats" which are heard when two notes of nearly equal pitch arc sounded together, and prove that the number of beats per second is equal to the difference in the frequencies of the notes.
9. If the velocity of sound in air is 1100 ft./sec., find the speed with which a whistle which gives a note of 256 vibrations per second must be moved so that to an observer at rest the pitch of the note heard is 258 vibrations per second.
10. Describe the stroboscopic disc method of measuring the pitch of a tuning-fork.
    If the disc makes 500 revolutions in 2 minutes when the ring containing 25 dots appears at rest, what is the pitch of the fork?
11. If the ring of dots in the above question in place of appearing at rest appears to slowly rotate in the same direction as that of rotation of the disc so that 20 dots appear to pass per minute, what would be the pitch of the fork?

CHAPTER IV

1. Given that the velocity with which a transverse wave travels along a stretched string is equal to $\sqrt{\frac{T}{m}}$, where T is the tension and m is the mass of unit length, show how to obtain an expression for the frequency of a string when giving its fundamental. If a wire a metre long gives a note of which the pitch is 200, what is the velocity of a transverse wave along a similar wire stretched with the same tension?
2. Two strings of the same length and material but of different diameters are stretched with the same tension. When caused to vibrate transversely the thicker string has a frequency double that of the other. Find the ratio of the diameters of the strings.
3. Given three wires of the same material and diameter, calculate the ratio of the lengths which when stretched with the same tension would give the notes C, E, and G respectively.
4. The tension T of a certain string is adjusted so that when vibrating with one loop, two loops, three loops, &c., the pitch of the note is always the same. Show that if N is the number of loops, then N^2T is a constant.
5. A string 2 metres long is stretched with a force equal to the weight of 10 kilograms. If the mass of a metre of the string is ·5 gram, what are the frequencies of the notes it is capable of emitting?
6. Explain why the notes which can be sounded by an organ-pipe depend on the length but not on the diameter. What are the frequencies of the notes which can be produced with a closed pipe 4 ft. long when the velocity of sound is 1100 ft./sec.?
7. What would be the effect on (1) the absolute pitch and (2) the relative pitch of the pipes of an organ if the temperature rose from 10° C. to 20°?

CHAPTERS V AND VI

1. Explain what is meant by resonance and give two examples of this phenomenon.
    Two pieces of the same wire are stretched side by side with the same tension, the length between the supports of one being twice as great as that of the other. If either wire is set into vibration by being plucked near the centre, what modes of vibration, if any, will be produced in the other?
2. A tall narrow cylindrical jar, 15 in. high, is found to resound to the note given by a fork. Taking the velocity of sound as 1100 ft./sec., calculate the pitch of the fork. Does the problem allow of your giving a definite answer. If not, state explicitly the conditions you assume in obtaining your result.
3. Given the velocity of sound in air, how would you determine the velocity in coal gas?
4. Explain the manner in which the vibrations in an ordinary organ-pipe are maintained.
5. How do you account for the fact that some intervals are concordant and others are discordant.

6. Explain why two notes of which the frequencies are 200 and 300 are concordant, but if the pitch of the lower is changed to 204 marked discord is produced.
7. Account for the characteristic of a vowel sound.

BOOK IV

CHAPTER I

1. Explain the difference between the nature of the shadows cast by a point source of light and by a source having appreciable size. If an observer move outwards from the centre of the shadow in each case, what will he see as he looks towards the source.
2. A candle is placed between two plane mirrors which are inclined at an angle of 90°. Draw a figure showing the position of the images, and trace from the candle to an eye looking into the mirrors the course of a pencil of rays by which each of the images is seen.
3. A square screen, each side 6 ft., has a small hole at its mid point, and a plane mirror is attached to a wall parallel to the screen. What is the smallest mirror which will enable a person looking through the hole to see the whole of the screen by reflection in the mirror.
4. Two plane mirrors are inclined at a fixed angle to one another, and a ray of light is reflected first in one mirror and then in the other, the plane of incidence being at right angles to the line of intersection of the mirrors. If the combination is rotated about this line, show that the direction of the reflected ray is unaltered.
5. Prove the relation between the radius of curvature of a concave mirror and the focal length, explaining carefully what is meant by radius of curvature and focal length.

   At what distance from a concave mirror of radius of curvature 10 cm. must an object be placed to form an image on a screen at 2 metres from the mirror ?
6. Describe an optical method by means of which the focal length of a concave mirror could be determined.
7. Determine the position of an object so that a virtual image twice the linear size of the object may be formed in a concave mirror of radius 10 cm.
8. Describe the sextant and prove that the angle as read on the instrument is half the angle between the objects sighted.

CHAPTER II

1. Give the laws of refraction of light and calculate the value of the critical angle for a substance of which the refractive index is 2.

   Describe in general terms the appearance presented to a fish by the bank of a stream below and above the surface of the water.
2. Describe what is meant by total internal reflection, and define the critical angle of a refracting substance in air. Find the greatest value of the refractive index of a specimen of glass if a corner of a rectangular block of it can act like a prism.
3. A point is viewed through a parallel plate of glass by an eye situated in the perpendicular from the point on to the glass. Show (1) by calculation, (2) by a carefully drawn figure, that the image will be nearer than the object by the same amount wherever the plate is.
4. A microscope is focussed on a scratch on a metal plate and then a slab of glass of refractive index 1·6 and thickness 2 cm. is placed over the plate. How far must the microscope be moved away for the scratch to again be in focus ?
5. The angle of a prism is 30° and a ray of light incident normally on one face is deviated through an angle of 15°. What is the refractive index of the prism ?

6. If a screen is placed at a fixed distance from an object, show that in general there are two positions in which a convex lens will form a sharp image on the screen.
7. Show that in the above arrangement the distance between the screens must be at least four times the focal length of the lens.
8. If in the above arrangement the height of the image in one position of the lens is 10 cm. and in the other position it is 44 cm., what is the height of the object?
9. An image of a small object is to be formed on a screen at a distance of 3 metres, and is to be magnified 5 times (linear). Find the focal length and position of the lens that will be required.
10. A convex lens and a convex mirror have a common axis. Rays from a small object on the axis pass through the lens, are reflected at the mirror, and form a real image coinciding with the object. Explain the conditions under which this occurs.
11. An empty vase is placed at a distance of 50 cm. from a convex lens of focal length 20 cm. How and where must a flower be arranged on the other side of the lens from the vase, so that to an eye suitably placed the flower may appear as if it were standing in the vase? Where must the eye be placed?
12. A parallel pencil of light traverses two co-axial convex lenses of focal lengths 10 and 5 cm. respectively. Draw diagrams representing the paths of the rays when the lenses are (1) in contact and separated by (2) 8 cm., (3) 15 cm., and (4) 20 cm.
13. The front lens of a camera is a converging lens of 10 in. focus, behind which at a distance of 6 in. is placed a diverging lens of 5 in. focus. Where will the image of a distant object be formed?

CHAPTERS III AND IV

1. You are given two convex lenses of focal lengths 20 cm. and 1 cm. respectively. How would you arrange them to form (*a*) a telescope and (*b*) a microscope?

   Draw in each case a diagram showing the course of the rays of light through the combination, and calculate the magnifying power of the telescope.
2. An astronomical telescope is formed by two convex lenses which are placed 50 cm. apart. If the linear magnification produced is 10, what are the focal lengths of the lenses?
3 Trace the course of rays through a telescope provided with (*a*) a converging lens and (*b*) a diverging lens for eyepiece. Show in each case where the eye should be placed to observe the image.
4. Explain, giving a diagram, the action of a simple magnifying glass.

   Will the magnifying power of such a glass be the same for a long-sighted person as for a short-sighted one?
5. The diameter of the sun subtends an angle of half a degree at the earth; what will be the size of the image formed by a convex lens of 2 metres focal length?
6. Give a general description of the structure of the eye.

   A short-sighted person can only see objects distinctly if their distance is between 5 and 20 inches. What spectacles must he use to see very distant objects, and what will be the distance of the nearest object he can see clearly when wearing them?
7. What is meant by the critical angle of a refracting medium?

   A pencil of light on its way to form an image is received normally on one-half of the hypotenuse of an isosceles right-angled prism, and emerges from the other half after two internal reflections. What is the effect on the image? If the emerging pencil is received similarly on a second prism like the first, but with the right-angled edge perpendicular to that of the first, show that the image is reversed.
8. Explain how you would project a spectrum on a screen.

   If this spectrum is viewed through the prism by which it is produced, what would be its appearance?

9. Explain why it is that the image of a small white source of light formed by a simple lens has a coloured edge, but that formed by a concave mirror is not coloured.
10. Define the dispersive power of a medium, and show that by means of two kinds of glass an achromatic lens can be constructed only if the dispersive powers of the glasses are different.

CHAPTER V

1. Describe Fizeau's method of measuring the velocity of light, and calculate the velocity if the distance between the wheel and the distant mirror is 8000 yards, and the first eclipse occurs when the wheel, which has 720 teeth, makes 13·75 revolutions per second.
2. Explain the principle of the method adopted by Römer to calculate the velocity of light from observations of the eclipses of Jupiter's satellite.

   If the satellite revolves round the planet in 40 hours, and the velocity of the earth in its orbit is 18 miles per second, find the greatest and least apparent intervals between successive eclipses, given that the velocity of light is 187,000 miles per second.
3. Give an account of the refraction of light by the aid of Huyghen's construction, and show that the refractive index from one medium to another is equal to the ratio of the velocities in the two media.
4. A drop of turpentine is placed on the surface of water in a tray. Explain the succession of colours observed as the turpentine evaporates. Describe the appearance of the spectrum of the light reflected from a slip of glass or mica which is just too thick to show colour with white light.
5. A slightly convex lens is placed on a plane glass plate. Explain the formation of the rings seen round the point of contact when it is looked at nearly normally.

   How could you show that the wave-length of red light is greater than the wave-length of blue light?
6. If a grating has 1000 lines to a centimetre, how far will the D lines (wave-length ·000060 cm.) in the first spectrum be from the central image on a screen 1 metre from the grating.
7. Describe Fresnell's biprism method of producing interference.

   If the distance between the slit and the screen is 100 cm., and that between the virtual images of the slit is half a millimetre, calculate the distance between the fringes produced with light of wave-length 6×10^{-5} cm.
8. Explain what is meant by the *equivalent air-path.*

   Apply the principle of equivalent air-path to show that when a parallel beam of light passes through a prism in such a manner that the ray inside the prism and the faces form an isosceles triangle, then the angles, the incident, and emergent rays made with the two faces of prism are equal.

CHAPTER VI

1. Explain what is meant by the candle-power of a source of light, and the intensity of illumination of a surface.

   Given a source of light of known candle-power, how would you determine the intensity of illumination on the surface of a table due to a given lamp?
2. A screen is illuminated by two 16 candle-power lamps placed at 2 and 5 metres respectively from it; at what distance must a 32 candle-power lamp be placed to give the same intensity of illumination?
3. Find a relation between the amount of light which falls on a given surface and its distance from the illuminating source.

   Describe the grease-spot photometer and explain how you would use it to verify the relation mentioned above.
4. A uniform white-hot surface is viewed through a black tube, and it looks equally bright at whatever distance it is from the end of the tube. Hence deduce the inverse-square law for the intensity of the illumination by a source

5. Distinguish between the candle-power and the intrinsic brightness of a source.

An electric glow-lamp has a candle-power of 16 and an incandescent gas mantle has a candle-power of 40; which has the greater intrinsic brightness?

6. A piece of platinum, 2 c.m. by 4 c.m., is heated by an electric current, and the candle-power in a direction at right angles to the plane is 10. A convex lens of 10 c.m. diameter is placed at 1 metre from the platinum and forms an image on a screen placed at 2 metres from the lens. Compare the intensity of illumination of the portion of the screen on which the image falls with what it would be if the lens were removed. Only 90 per cent. of the light incident on the lens is transmitted, the rest being reflected at the surfaces.

CHAPTER VII

1. Explain how the distribution of energy in the spectrum varies with the temperature of the source.
2. Describe some form of radiation pyrometer, and explain the principle on which it works.
3. How would you show that a black surface has higher radiating power and higher absorbing power than a polished metal surface?

   Part of a platinum wire has its surface lightly covered by lamp-black, and an electric current is sent through the wire. Will the blackened portion be hotter or cooler than the rest?
4. Two mercury thermometers, one having a blackened bulb and the other a silvered bulb, are placed close together (1) in the sunlight, and (2) in a closed box, the walls of which are at a constant temperature. Will the readings of the thermometers be the same? Give reasons for your answer.
5. Explain why it is that a mixture of yellow and blue pigments appears green.

CHAPTER VIII

1. Describe an experiment which indicates that light is propagated by a transverse wave motion.
2. Explain what is meant by the plane of polarisation, optic axis, principal plane, ordinary ray, and extraordinary ray, giving any connection which may exist between these various quantities.
3. Describe the construction of a Nicol prism, and explain how you would determine the plane of polarisation of the light passing through.
4. Describe an experiment by means of which you would investigate the question as to whether a given sample of glass was in a state of strain.
5. Explain what is meant by the polarising angle, and show that if this angle is given by the relation $\tan \theta = n$, where n is the refractive index, then when light is incident at the polarising angle the reflected and transmitted beams are at right angles to one another.

BOOK V

CHAPTERS I AND II

1. Prove that the intensity of the magnetic field due to a short bar-magnet at a point on the axis produced is approximately twice the intensity at a point the same distance from the centre of the magnet, but on a line at right angles to its length.
2. Find an expression for the couple acting on a magnet when placed in a uniform magnetic field, and show how the expression may be used to define the magnetic moment of a magnet.

   Calculate the work which must be performed to twist a magnet of which the moment is 250 C.G.S. in a horizontal plane through an angle of 60° from the magnetic meridian, the value of H being ·18 C.G.S.

3. A short bar-magnet is placed with its axis at right angles to the magnetic meridian, and a compass-needle is at a distance of 20 cm. from the centre and due E. of the magnet. If as a result the compass points magnetic NW., what inferences would you draw as to the position and strength of the magnet? ($H = \cdot 2$ C.G.S.).
4. Find an expression for the period of a bar-magnet which is suspended by a fine thread in a magnetic field so that its axis is horizontal.

   If such a magnet makes 100 vibrations at A in one minute and 120 vibrations at B, find the ratio of the value of H in these two places.
5. A bar-magnet is placed horizontally in the magnetic meridian with its S. pole towards the north, and at a point distant 10 cm. from the S. pole the resultant field strength is zero. If the distance between the poles of the magnet is 10 cm. and the value of the horizontal field due to the earth is ·2, find the pole strength of the magnet.
6. A short bar-magnet, moment 500 C.G.S. units, lies with its axis horizontal and pointing magnetic north. A small magnetic needle, 20 cm. north of the centre of the magnet, makes 20 vibrations per minute. When the magnet is turned through 180° about its centre the needle makes 8·49 vibrations per minute. Find the value of H.
7. A and B are the poles of a thin magnet, the distance between the poles being 20 cm. and the strength of each pole is 10 units. Find the magnitude and direction of the magnetic field at a point C which is equi-distant from A and B.

   In the above question, what would be the force exerted upon the magnet by a pole of strength of 20 units placed at C?
8. Explain why it is that in general a magnetised needle which can turn freely about a horizontal axis through its centre of gravity will set itself vertical if the axis is in the magnetic meridian.

   What would occur at the magnetic equator?
9. Describe the experiments necessary to determine the horizontal component of the earth's magnetic field, proving the formulæ used in obtaining the result from the observed quantities.
10. Give a general description of the way in which the earth's magnetic field varies at different points on the earth's surface.
11. A very long thin magnet is supported in a vertical position with its north pole on the surface of a table, and a circle is described with this pole as centre and passing through the point where the field due to the magnet is exactly equal and opposite to the horizontal field of the earth. What is the direction and magnitude of the resultant field at points on this circle magnetic N., W., S., and E. of the pole?

CHAPTER III

1. A steel rod is magnetised longitudinally and is then placed inside an iron tube of equal length, the inside of the tube being of slightly greater diameter than that of the rod. What will be the effect of the presence of the tube on the field due to the magnet? Draw diagrams illustrating the distribution of the lines of force of the magnet before and after the tube is placed in position.
2. Explain what is meant by the intensity of magnetisation.

   A cylinder of steel 1 cm. in diameter and 50 cm. long is magnetised so that its moment is 5000 C.G.S. Assuming that the poles are at the end, what is the intensity of magnetisation of the steel near the middle of the bar?
3. Explain carefully the meaning of each of the terms in the following equation—$B = H + I4\pi$—and give the names by which the ratios of the different quantities are known.
4. Give a general account of the manner in which the induction through soft iron changes as the magnetising force is changed.
5. Give a general account of the effect of temperature on the permeability of iron.
6. What is meant by hysteresis, and how can it be explained on the hypothesis of molecular magnets?

CHAPTER IV

1. Two small conducting spheres, each weighing 0·1 gram, are suspended by silk fibres 1 metre long, the upper ends being at the same level and 5 cm. apart. What charge must be communicated to each sphere so that they may come to rest at a distance of 3 cm. apart?
2. A condenser is formed by two parallel co-axial circular discs of 20 cm. radius, 1 mm. apart. Neglecting the effect of the edges, calculate the capacity when the dielectric is air.
   What is the energy of the charge when the charge on the insulated plate is 100 electrostatic units?
3. Give an account of the Leyden jar, and state on what the charge it receives will depend.
   If you are given two Leyden jars, a source which will charge them at given potential, and a gold-leaf electroscope, how can you find which has the greater capacity?
4. Find the work necessary to withdraw the plates of an air condenser through 0·2 cm. if charged with 600 electrostatic units of electricity. The plates are discs 30 cm. in diameter, and are 1 cm. apart. What becomes of the work done?
5. Explain what is meant by the capacity of a condenser.
   The capacities of two Leyden jars are C_1 and C_2; compare the energies of their charges under the following conditions: (*a*) when their charges are equal, and (*b*) when the difference of potential between the coatings is the same.
6. Two insulated metal spheres of radii 5 and 7 cm. are placed at a considerable distance apart and charged positively with 10 and 6 units respectively. What will be the charges on the spheres after they have been connected by a fine insulated wire?
7. If the spheres in the above question are brought near together, will their respective electrical states be altered, and if so, in what direction?
8. A hollow insulated conductor is electrified positively. A gold-leaf electroscope is electrified negatively, and a small conducting sphere connected by a thin wire to the electroscope is (1) brought near the conductor, (2) introduced, and (3) made to touch the inside. Describe and explain the indications of the electroscope.
9. Find an expression for the energy of a charged condenser, defining carefully what is meant by the quantities in terms of which you express the energy.
10. Show that the capacity of a condenser consisting of two concentric spherical shells of nearly equal radius is equal to $SK/4\pi d$, where S is the surface of either shell, d the difference in their radii, and K the specific inductive capacity of the medium between the shells.
11. Describe a form of attracted disc electrometer.
    If the movable disc has a diameter of 5 cm. and the distance apart of the discs 0·5 is cm., calculate the P.D. when the pull on the movable disc is 800 dynes.
12. Explain what is meant by the potential and the electric intensity at a point in the air. How would you actually determine the potential at a given point in the air?

CHAPTER V

1. State Ohm's law and explain its meaning.
   A galvanometer has a resistance of 10 ohms, and gives a deflection of one scale division when traversed by a current of 10^{-6} amperes. What resistance must be placed in series with the galvanometer so that it may give a deflection of 100 divisions when a P.D. of 100 volts is applied?
2. The E.M.F. of a battery is 4 volts and its internal resistance is 1·5 ohms. If the poles are joined by a wire of which the resistance is 20 ohms, calculate the current produced and the difference of potential between the ends of the wire.

3. The terminals of a galvanometer are connected by a wire which has the same resistance as that of the galvanometer, and a lead from a battery is connected to one end of this wire. Find the point on the wire at which the other lead from the battery has to be connected so that exactly one-tenth of the current supplied by the battery will traverse the galvanometer.
4. A battery of which the E.M.F. is 4 volts and the internal resistance is 0·2 ohm has its terminals joined by two wires of resistances 10 and 5 ohms placed in parallel. Find the current in each wire and the amount of heat in ergs generated per second in each.
5. Show how by means of an accurately calibrated ammeter and voltmeter we can determine the mechanical equivalent of heat. Give a diagram of apparatus and connections.

   A wire is immersed in 100 grams of water in a calorimeter. Find the rise of temperature of the water in 5 minutes when a current of 5 amperes passes through the wire whose resistance is 3 ohms. ($J = 4{\cdot}2 \times 10^7$ C.G.S. units.)
6. A battery has an E.M.F. of 18 volts and an internal resistance of 3 ohms. The P.D. across the terminals when connected by a wire A is 15 volts, when by a wire B 12 volts. Compare the heat developed in A and B in equal time.
7. Explain what is meant by the term conductance, and show that the conductance of two conductors placed in parallel is equal to the sum of their conductances.
8. If the resistivity of platinum is 8×10^{-6} ohms per cm. per sq. cm., calculate the length of a wire of diameter ·2 mm. which will have a resistance of 1 ohm.
9. The P.D. of an electric supply is 100 volts; what must be the resistance of an electric radiator so that when connected to this supply it may give out as much heat as is obtained by burning 10 lbs. of coal per hour, the calorific value of the coal being $3{\cdot}6 \times 10^6$ *calories* per *lb.?*
10. How many kilowatt-hours of electrical energy will the above radiator require if it is run 6 hours a day for a week?

CHAPTER VI

1. Define the unit of current in electro-magnetic measure, and state what is 1 ampere.

   A coil of 10 turns and 10 cm. radius is placed in the magnetic meridian with a needle at its centre. Find the current in amperes which will deflect the needle 45° if the horizontal intensity is 0·16.
2. Deduce a simple expression for the work done when a magnetic pole passes along any closed path round a current, and hence deduce the intensity of the magnetic field inside a long solenoid.
3. A current of 5 amperes is passed through a coil containing 20 turns of mean radius 15 cm.; find the strength of the magnetic field in C.G.S. units at the centre of the coil.
4. Describe and explain the use of a tangent galvanometer.

   If the coil of a tangent galvanometer has 20 turns, and a current of 9 amperes produces a deflection of 45° at a place where $H = {\cdot}18$ C.G.S. units, what is the mean radius of the coil?
5. Give Ampère's expression for the magnetic field due to an element of current, and deduce the intensity of the field at any point on the axis of a circular wire carrying a current.
6. Explain what is meant by the terms magnetomotive force and reluctance.
7. Find an expression for the field inside an anchor-ring which is uniformly wound with a layer of insulated wire through which a current is passed, and discuss the question as to whether this field is uniform.
8. A straight wire rolls on two parallel horizontal cylinders 2 metres apart. If a current of 20 amperes is passed from one cylinder through the wire to the other cylinder, find the force which must be applied to the wire to prevent it moving. The vertical component of the earth's field is ·4 C.G.S.

CHAPTER VII

1. Explain the action of an earth inductor and describe how it could be used to determine the magnetic dip.
2. A metal rod a metre long is rotated about one end as a centre so as to sweep out a circle of a metre radius at right angles to a uniform magnetic field, of which the strength is 1000 C.G.S. units. How many revolutions per second will be required to produce a potential difference of a volt between the ends of the rod?
3. Two long insulated metal rods are fixed parallel to each other on a table, and one end of each is connected to a galvanometer. A metal cylinder rolls on the rods, and as a result the galvanometer is deflected. Explain this deflection, and draw a diagram showing the connection between the direction in which the cylinder moves and the direction of the current produced. On what factors does the magnitude of the current in the galvanometer depend?
4. Two coils of wire are placed close together, one being connected to a battery and the other to a galvanometer. Describe and explain the effect produced on the galvanometer when (1) the current is started, (2) the current is stopped, (3) the coils are moved nearer together, and (4) are moved further apart.
5. A copper hoop is set into rotation about a diameter as axis in a magnetic field with the axis (1) parallel to the field and (2) perpendicular to the field. If the friction is the same in the two cases and the initial velocity is also the same, explain why the hoop comes to rest more quickly in one case than in the other.
6. Explain what is meant by the terms self and mutual inductance.
7. An anchor-ring wound on a wooden core of 2 sq. cm. cross-section, the mean radius of the core being 20 cm., contains 500 turns of insulated wire. What is approximately its self-inductance?
8. Describe the Gramme armature and the manner in which it works.
9. Describe and explain, illustrating your answer by a diagram, one method of winding a drum armature. What is the advantage of the drum winding over the Gramme winding?
10. A motor is worked by a battery of accumulators. Show that the fraction of the energy put out by the battery available for conversion into work decreases as the current rises.
11. A shunt-wound motor is running light; explain why the speed of the motor increases if the resistance in the field-coil circuit is increased.
12. What power is expended in turning the armature of a series motor if the P.D. of the supply is 100 volts, the resistance of the field coils and armature is ·6 ohm, and the current taken is 20 amperes? What is the electrical efficiency of the motor?
13. Explain what is meant by the root-mean-square of an alternating current.
    If the R.M.S. value of a current is 250 amperes, what is the greatest instantaneous value of the current?
14. Show that two equal alternating magnetic fields in directions at right angles are equivalent to a uniform field rotating at a uniform rate.
15. Explain the principle of the transformer.
    Why is it that very little power is absorbed in the primary of a transformer when the secondary is on open circuit?
16. Describe and explain an arrangement for sending messages simultaneously in both directions through a cable.

CHAPTER VIII

1. Give the laws of electrolysis, and explain what is meant by the electro-chemical equivalent.
2. A current is passed through three electrolytic cells placed in series, containing (1) a solution of sulphuric acid, (2) a solution of copper sulphate, and (3) a solution of silver nitrate. How much hydrogen, oxygen, and copper will be

liberated during the time 10 grams of silver is deposited, given the following chemical equivalents: H=1, O=16, Cu=32, Ag=108.

3. A copper voltameter is connected in series with a wire AB and an accumulator, and a deposit of 1·677 grams of copper is obtained in 1 hour. Another wire CD is connected to a constant source of potential E. When A is connected to C, and at the same time B is joined through a galvanometer to a point on CD, 142 cm. from C, no deflection is obtained in the galvanometer. Find the resistance of the wire AB, given that E is a 2-volt accumulator of negligible resistance, and the total length of CD is 400 cm. Electro-chemical equivalent of copper = ·00328 *C.G.S. units.*
4. Give a general explanation of the conduction of electricity through an electrolyte. Find the time necessary to decompose 1 kilogram of water by a current of 10 amperes if the separation of each gram of hydrogen requires 96,500 coulombs.
5. Discuss the question as to whether the E.M.F. of a Daniell cell depends on the dimensions of the metal plates and their distance apart. In what electrical respects does a large cell differ from a small one?
6. Explain what is meant by polarisation in electrolysis.
7. Describe the construction of a lead storage cell (accumulator) and give the chemical changes which take place during charge and discharge.

CHAPTER IX

1. What evidence is there for the hypothesis that the cathode rays consist of negatively charged particles, and what approximately is the mass of these particles?
2. Explain what is meant by the saturation current in a gas, and describe how this quantity can be determined.
3. Describe the properties of the rays given off by radium.
4. Give a short account of the disintegration hypothesis in the case of radio-active substances.

CHAPTERS X AND XI

1. Describe some method by which electro-magnetic waves may be produced.
2. What is the arrangement employed to send out and receive wireless telegraphic messages?
3. The ends of two copper wires connected to a galvanometer are joined to an iron wire. One of the junctions is kept at 0°, and the other is heated to a bright red and then allowed to cool slowly. Describe and account for the manner in which the deflection of the galvanometer alters during the cooling.
4. Describe how a thermo-electric junction may be employed to measure a temperature.

ANSWERS TO THE NUMERICAL QUESTIONS

(In most cases the last significant figure given is only correct to within two or three units in this place.)

BOOK I

CHAPTERS I, II, AND III

1. 571 cm. **2.** 5·66 hours. **4.** ·1117 ft./sec.^{2}. **5.** 250 dynes. **6.** 500 seconds. **7.** 6 ft. from A. **8.** (*a*) 137·2 lbs. and 22·8 lbs.; (*b*) 100 lbs. **9.** 13860 dynes = 1 poundal. **11.** 7167. **12.** 3·75 lbs. weight. **13.** 4×10^5 ft.-lbs.; 1·21 H.P. **14.** ·727 H.P. **15.** 3466 ft.-lbs. **16.** K.E., 384 and 640 foot-poundals. Tension, 3·75 lbs. weight.

CHAPTERS IV, V, AND VI

2. 25·7 cm. **3.** $2\frac{2}{9}$ cm. from centre. **4.** 5089 seconds. **5.** $5·55 \times 10^5$ ft.-lbs. **6.** 366·5 lb.-ft. **10.** 22·85 H.P. **12.** 112 lbs. per ton. **13.** 6 H.P., 12 H.P.

CHAPTERS VII AND VIII

1. $4·9 \times 10^6$ dynes per sq. cm. **3.** 31·25 in. **6.** 204 in. **7.** 25 cubic ft. **9.** 69·37 in. of mercury.

CHAPTERS IX, X, AND XI

1. Vol.=2 c.c., density=1·5. **2.** 62,420 lbs. **3.** ·11. **4.** 2·5 in. **6.** 955 lbs. **7.** Vol. of A is ·4 of vol. of B. **10.** 25,120 ergs. **11.** 8·16 c.m. **12.** 8014 grams. **15.** 41·667 kilograms applied at lower edge.

BOOK II

CHAPTER I

1. 100·044 cm. **2.** 2·23 mm. to be subtracted. **7.** ·00366. **8.** ·000182. **9.** ·000465. **10.** ·001103. **11.** $2·86 \times 10^6$.

CHAPTERS II AND III

1. 25 grams of ice and 125 grams of water, all at 0° C. **3.** 25·6 grams of ice, all at 0°. **4.** ·562. **6.** 19·77°. **9.** ·122. **18.** 36·4 per cent.

CHAPTER IV

4. 1·56 grams per second. **5.** $3·24 \times 10^9$ calories. **6.** 600 grams. **7.** 2·38 calories per second.

CHAPTER V

1. $8·198 \times 10^5$ cm./sec. **2.** 152 metres/sec. **3.** 0·75° C. **12.** ·423.

BOOK III

CHAPTERS I AND II

6. 314 cm./sec., 97400 cm./sec.^{2} **9.** ·777. **11.** Iron rods 2·5 times length of zinc rods. **14.** $2·263 \times 10^6$, $4·527 \times 10^5$, $3·394 \times 10^5$ dynes.

CHAPTER III

1. 546·4° C. **2.** $1·85 \times 10^{12}$. **4.** 1·427. **6.** 66·4 cm./sec. **9.** 8·59 ft./sec. towards observer. **10.** 52·13. **11.** 51·96.

CHAPTER IV

1. 400 metres/sec. **2.** 2 : 1. **3.** 6 : 5 : 4. **5.** 22·66, 45·32, 67·98, &c. **6.** 68·76, 206·28, 343·80, &c. **7.** $\frac{10183}{10366}$.

CHAPTERS V AND VI

2. 220.

BOOK IV

CHAPTER I

3. 3 ft. × 3 ft. **5.** 5·13 cm. **7.** 2·5 cm. from mirror.

CHAPTER II

1. 30°. **2.** 1·41. **4.** ·75 cm. **5.** 1·41. **8.** 20·98 cm. **9.** 41·67 cm. lens at 50 cm. from object. **11.** Inverted and at 33·33 cm. from lens. **13.** 20 in. beyond concave lens.

CHAPTER III AND IV

2. 45·46 and 4·54 cm. **5.** 1·745 cm. in diameter. **6.** Concave, focal length 20 in., 6 67 in.

CHAPTER V

1. $1{\cdot}8 \times 10^5$ miles/sec. **2.** 40 hours ± 13·86 seconds. **6.** 6 cm. **7.** 1·2 mm.

CHAPTER VI

2. 83 cm. **6.** 19·89.

BOOK V

CHAPTERS I AND II

2. 22·5 ergs. **3.** Moment 800 C.G.S. with S. pole towards E. **4.** Field at B : field at A : : 1·41 : 1. **5.** 133·3 C.G.S. **6.** ·18 C.G.S. **7.** ·025 C.G.S. The force on the magnet is a couple of 8·66 dynes-cm. and a force of ·5 dyne parallel to the length of the magnet. **11.** At N. point $2H$ in northerly direction ; at W. point $\sqrt{2}.H$ in NE. direction ; at S. point zero ; at W. point $\sqrt{2}.H$ in NW. direction.

CHAPTER III

2. 127·4 C.G.S.

CHAPTER IV

1. + and − 2·97 electrostatic units. **2.** 10 cm. 500 ergs. **4.** 6·4 ergs. **5.** (*a*) C_2/C_1, (*b*) C_1/C_2. **6.** 6·67 and 9·33 units. **11.** 16 electrostatic units.

CHAPTER V

1. 999,990 ohms. **2.** 1·865 amperes, 37·3 volts. **3.** So as to include a fifth of the total length of the wire. **4.** ·378 and ·756 amperes ; $1{\cdot}43 \times 10^7$ and $2{\cdot}86 \times 10^7$ ergs. **5.** 56°·2. **6.** 15 : 24. **8.** 39·27 cm. **9.** ·238 ohm. **10.** 1764.

CHAPTER VI

1. ·267 ampere. **3.** 4·19 C.G.S. **4.** 10 cm. **8.** 160 dynes.

CHAPTER VII

2. 3·18 revolutions per second. **7.** 50,000 cm. **12.** 1760 watts, ·88. **13.** ± 354 amperes.

CHAPTER VIII

2. H = ·093, O = 1·48, Cu = 2·96 grams. **3.** $\frac{1}{2}$ ohm. **4.** 12 days, 9 hours, 50 minutes, 4 seconds.

INDEX

THE END

Printed by BALLANTYNE, HANSON & Co.
Edinburgh & London